DOCUMENT RESUME

ED 412 249 TM 027 517

AUTHOR	Hartka, Liz; Stancavage, Fran
TITLE	Perspectives on the Impact of the 1994 Trial State Assessments: State Assessment Directors, State Mathematics Specialists, and State Reading Specialists.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, CA.
SPONS AGENCY	National Academy of Education, Washington, DC.
PUB DATE	1997-00-00
NOTE	56p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Administrator Attitudes; Administrators; Case Studies; *Grade 4; Intermediate Grades; *Mathematics Tests; National Surveys; *Reading Tests; Test Construction; Test Interpretation; Test Results; *Test Use
IDENTIFIERS	Impact Evaluation; National Assessment of Educational Progress; *Test Directors; *Trial State Assessment (NAEP)

ABSTRACT

This paper is the fourth in a series prepared for the National Academy of Education on the evaluation of the National Assessment of Educational Progress Trial State Assessment (TSA) and the impact of reporting TSA results. This paper provides a perspective on the last of the TSAs, an assessment of fourth-grade reading that was carried out in 44 participating states and territories in February 1994. Results of this TSA were released in a preliminary version in April 1995 and in reports in August and October 1995 and March 1996. A questionnaire about the overall impact of the 1990, 1992, and 1994 TSA assessments was sent to assessment directors and curriculum specialists in all 50 states and the District of Columbia, and 9 case study interviews (in 9 states) were carried out to explore the impact of TSA results. The overall impact of the TSAs has been viewed as generally positive, with about half the states evaluating the TSAs positively, and none evaluating them as having negative overall impact. The case studies made it clear that states are the primary consumers of TSA information, and that the impact of the TSAs, including the 1994 TSA, was greatest in states in which performance was worst. For the 1994 TSA, the impact of the reading assessment seems to have been mediated by the extent to which instruction is subject to local control. Some weaknesses of the TSA program and the 1994 TSA in particular were identified, but overall, the National Assessment of Educational Progress appears to have sustained its perceived value to educators and policymakers in its TSAs. An appendix presents the nine case studies. (Contains 1 figure and 18 tables.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 249

Perspectives on the Impact of the 1994 Trial State Assessments: State Assessment Directors, State Mathematics Specialists, and State Reading Specialists

Liz Hartka
Fran Stancavage
American Institutes for Research

TM027517

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Fran Stancavage
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

♦ Introduction

This is the fourth in a series of research papers prepared for the National Academy of Education (NAE) Panel on the Evaluation of the National Assessment of Educational Progress (NAEP) Trial State Assessment (TSA) that have examined the impact of reporting TSA results.

1990 TSA

The first paper, released in January 1992, provided a look at the immediate impact of reporting results from the first TSA: the 1990 assessment of eighth-grade mathematics.[1] Based on telephone surveys conducted shortly before and after the June 1991 release, it concluded that the results of the assessment had been widely disseminated and had given rise to meaningful discussions among a number of groups concerned with education policy. There were indications that results were beginning to influence state-level changes in mathematics instruction and assessment within weeks after their release.

A second paper continued the evaluation of the impact by examining the longer term influences of the 1990 results.[2] The main data sources for the study were a telephone survey of national and state respondents carried out in the first months of 1992, and a set of in-depth case studies completed in the fall of that year. Results indicated that, while there had been little evidence of penetration below the level of state legislatures and state departments of education, the majority of stakeholders at these levels reported a positive impact on education in their states and an expectation that impact would increase once the results of the 1992 TSA were released.

The second paper further concluded that the 1990 results carried weight, not in isolation, but because they articulated well with other contemporary influences in mathematics education, particularly the National Council of Teachers of Mathematics (NCTM) standards. Over 40 percent of state department of education respondents reported that curricular or assessment reforms in their states were being influenced by the NAEP/TSA. These respondents credited the NAEP/TSA with adding impetus to already planned or desired changes, accelerating discussions of state goals, and increasing pressure to align with national standards. Some very specific uses and influences attributed to the TSA included tipping the balance in favor of calculators (in the classroom and on assessments) and using sample NAEP items as models for states' own assessments and for purposes of teacher in-service training. Some respondents, however, also pointed out that NAEP's relevance for local schools or districts was diminished by competition with states' own assessments, particularly in those states where the state assessment was closely aligned with classroom practice. In general, it appeared that the TSA's impact would likely continue to be mediated through the state departments of education or through legislatively mandated changes in state assessments or teacher requirements, and that the impact of NAEP would be inversely related to the investment a state had made in its own assessment system.

1992 TSA

---

[1] F.B. Stancavage, E.D. Roeber, and G.W. Bohrnstedt, "A Study of the Impact of Reporting the Results of the 1990 Trial State Assessment: First Report," in *Assessing Student Achievement in the States: Background Studies* (Stanford, CA: The National Academy of Education, 1992).

[2] F.B. Stancavage, E.D. Roeber, and G.W. Bohrnstedt, "Impact of the 1990 Trial State Assessment: A Follow up Study," in *The Trial State Assessment: Prospects and Realities: Background Studies* (Stanford, CA: The National Academy of Education, 1993).

The third paper summarized the perceptions of state-level respondents regarding the impact of the 1992 TSAs in reading and mathematics.[3] State assessment directors and mathematics specialists were surveyed in the summer of 1993, shortly after the release of the mathematics results; assessment directors were surveyed again, along with reading specialists, in early 1994. The latter data collection was timed approximately 5 months after the release of the reading results.

Once again, the majority of respondents reported that participation in the TSA was a worthwhile endeavor and that the 1992 TSA had had a positive impact on their states' own instructional and assessment programs, albeit a relatively minor one. Those who saw the exercise as worthwhile explained that the TSA's value derived from the comparison it allowed between the states and the nation, and from the impetus it provided for change. The smaller percentage who indicated that participation was of limited or no value noted that their responses were made in the context of tight budgets and competing priorities, and they pointed to the absence of specific linkages between NAEP and states' own curricular goals as well as to the fact that NAEP does not provide results below the state level.

1994 TSA

The present report provides a perspective on the last of the TSAs, an assessment of fourth grade reading which was carried out in 44 participating states and territories during February, 1994.[4] Figure 1 shows the time line for the release of the 1994 TSA results.

Figure 1 — *Time line for the release of the 1994 TSA results in fourth-grade reading*

Date	NAEP Results
April, 1995	Release of *NAEP 1994 Reading: A First Look* — summary and highlights of national and state results
August, 1995	Data errors uncovered and plans for reanalysis of 1992 and 1994 reading results announced
October, 1995	Release of corrected version of *NAEP 1994 Reading: A First Look*
March, 1996	Release of State reports and *NAEP 1994 Reading Report Card for the Nation and the States*

In an effort to be responsive to requests for more timely data release and shorter, more user-friendly reports, the National Center for Education Statistics (NCES) released the core results of the 1994 reading TSA in stages. The initial offering was the *First Look* report, which provided highlights of the

[3]E. Hawkins, "Impact of the 1992 Trial State Assessment," in *Quality and Utility: the 1994 Trials State Assessment in Reading, Background Studies* (Stanford, CA: The National Academy of Education, 1996).

[4]After 1994, state NAEP assessments moved to a new status; the statute presently considers them to be "developmental," but no longer a short-term "trial."

1994 state and national reading results and was released in April, 1995. The report format was considered successful, but the goal of more rapid reporting was not met, given that the results once again lagged behind administration by about 13 months.

Release of more comprehensive reports was further delayed when, in August, 1995, scientists at Educational Testing Service (ETS) discovered an documentation error in the ETS version of the PARSCALE program, which is used to compute NAEP scale score results. Assessment results for the 1992 and 1994 national and state reading assessments and the 1992 national and state mathematics assessments were affected by this error. At about the same time, an additional error was discovered in the procedures used by American College Testing (ACT) in 1992 to translate the reading achievement levels into cut points on the NAEP scales. (The procedures contained an incorrectly derived formula.) NCES and the National Assessment Governing Board (NAGB) instructed ETS and ACT to immediately calculated revised reading results and publish a corrected version of the reading *First Look* report.[5] The revised results were not substantively different in most respects, and the rerelease, which was made available in October, 1995, was generally ignored by the media.

The NAE Evaluation Panel had planned to defer data collection on the impact of the 1994 TSA until after the release of the more comprehensive *1994 Reading Report Card for the Nation and the States* and the accompanying individual state reports. Partly in consequence of the delay occasioned by the reanalysis, however, these latter reports were not released until March 1996, after the data collection phase of the NAE evaluation had drawn to a close. Unable to await this event, the Panel commissioned a final survey of impact for December 1995. This mail survey, and a complementary set of case studies which were also conducted in December, focused on the *overall* impact of the TSA program since 1990.

This Report

This report summarizes the results of the 1995 impact study data collections and is organized around the following research questions:

- What has been the perceived overall impact of the TSA on education in the states?
- To what extent has the TSA influenced state instructional or assessment practices in reading or mathematics?
- What specific contextual factors influenced the impact of the TSA results in individual states?
- How highly do consumers value NAEP as a monitor of education programs?
- What are perceived to be the TSA's major weaknesses?
- Has participation in NAEP been viewed as a worthwhile exercise? Do states plan to participate in the future?

♦ Methodology

There were two primary data collections for this report:

- First, a brief paper-and-pencil questionnaire was designed to elicit opinions about the overall impact of the 1990, 1992, and 1994 TSA assessments. The questionnaire was distributed to assessment directors, reading curriculum specialists, and mathematics

[5] Mazzeo, J. (October 6, 1995) *The Network News*. Princeton, NJ: Educational Testing Service.

curriculum specialists at state departments of education in all fifty states plus the District of Columbia.

- Secondly, a series of nine case study interviews was carried out to provide detailed descriptions of the impact of the TSA in selected states as well as insight into the reasons why those impacts took place. Lessons from the case studies are summarized in the body of this report, and the full text of the case studies is included as an appendix.

Additional information is drawn from a survey of State Testing Directors that was administered in May, 1994 for the Study of the Administration of the 1994 TSA.[6] This survey focused on how well the TSA was administered in the states, states' expectations of the TSA program, whether those expectations were met, and observations on the future of the TSA program.

♦ Survey on Impact of Reporting

State assessment directors, mathematics specialists, and reading specialists from all 50 states and the District of Columbia (153 persons) were invited to respond to a brief mail survey during the months of December, 1995 and January, 1996. The survey was designed to elicit information about *the types of changes* that had occurred in reading and mathematics instruction and assessment in respondents' states since 1990 (when the TSA program was instituted). It was also intended to gather respondents' opinions about 1) the *influence* of the TSA on the aforementioned changes, 2) the *limitations* of the TSA, and 3) the *overall influence* of the TSA program.

Response rates for assessment directors, mathematics and reading specialists, broken out by states that had participated in at least one TSA (in 1990, 1992, or 1994) versus states that had never participated, are shown in table 1. Here and elsewhere in this report, the District of Columbia is counted as a state.

Table 1 — *Response rates for 1995 survey by respondent groups, separately for participating and non-participating states*

Participation Status	Respondent Group	Response Rate
Participated in at least one of the 1990, 1992, or 1994 TSAs (N = 46)	Assessment directors	89
	Mathematics specialists	80
	Reading specialists	83
Never participated in the TSA (N = 5)	Assessment directors	60
	Mathematics specialists	60
	Reading specialists	60

[6]L. Hartka, J. Yu, and D. McLaughlin, "A Study of the Administration of the 1994 Trial State Assessment," in *Quality and Utility: the 1994 Trials State Assessment in Reading, Background Studies* (Stanford, CA: The National Academy of Education, 1996).

Case Study Analysis

To complement the mail survey, AIR undertook a series of case studies designed to place the impact of the TSA within the broader context of each state's unique needs and circumstances. Case study states were chosen to represent all parts of the United States and to include both large and small-population states. In addition, preference was given to states where it appeared, on the basis responses to earlier rounds of Impact Study surveys, that we might find positive, measurable impact of the TSA on education.

An earlier set of case studies, carried out in fall, 1992, had been based on interviews with state assessment directors and mathematics specialists and focused on the impact of the TSA in mathematics. For the present case studies, researchers at AIR conducted semi-structured telephone interviews with assessment directors and reading specialists, and there was correspondingly greater emphasis on the impact of the TSA on reading. Respondents reviewed and approved the resulting case study reports, which are included here as appendix A.

The following nine states participated in the case studies:

- Connecticut
- Hawaii
- Louisiana
- North Carolina
- Pennsylvania
- Rhode Island
- West Virginia
- Wisconsin
- Wyoming

Two other states were approached for case studies but declined to participate either because the incumbent assessment director lacked experience with the TSA or because of competing state priorities.

♦ Research Findings

Results from the mail survey, case studies, and ancillary data sources are used below to answer research questions regarding perceptions of the impact of the TSA, the weaknesses of the TSA, and the value of state participation in the TSA.

What has been the perceived overall impact of the TSA on education in the states?

In the decade leading up to the first TSA, opinions about the value of a state NAEP program were mixed. Although a majority of chief state school officers eventually came out in favor of a state-NAEP trial, concern over the kinds of impact that might rebound upon the states remained. Some stakeholders, for example, were fearful of unwarranted federal influence on state education goals, and others thought that the "horse race"engendered by the state-to-state comparisons would bring pressure to bear on the wrong aspects of education reform.

The fact that these feared consequences did not materialize has been shown by the Panel's previous impact studies and is reinforced by the present findings. *Results in table 2 indicate that, among states that had participated in the TSA program at least once since 1990, the TSA's overall impact has been viewed as*

generally positive. About half the state assessment directors and a third each of the reading and mathematics specialists in these states evaluated the overall impact of the TSA program on education as being generally positive. None evaluated it as generally negative, and the remainder either had no opinion or evaluated the impact as mixed or too limited to classify.

Table 2 — ***Evaluative judgment of overall impact of the TSA program on state education, among states that participated in at least one TSA***

	Percent of Assessment Directors	Percent of Reading Specialists[1]	Percent of Math Specialists[2]
Generally positive	48	33	31
Generally negative	0	0	0
Mixed	8	33	8
Too limited to classify	40	22	44
Don't know/other	5	11	17
Number Responding	40	36	36

Source: Overall, which best describes the impact of the NAEP state assessments to date on education in your state? [1995 Impact Study Questionnaire]
[1]Restricted to reading specialists from states that participated in at least one TSA in reading.
[2]Restricted to mathematics specialists from states that participated in at least one TSA in mathematics.

Compared to the assessment directors and mathematics specialists, a relatively high proportion of the reading specialists, about one third, reported that the TSA's impact had been "mixed" in their states. Based on the case studies, we inferred that many of the reading specialists who chose this option did so in the belief that the impact of the TSAs was dependent upon the style of reading instruction endorsed by local educators in their states.

As of the end of 1995, most state level reading practices were aligned with—or moving in the direction of—literature based, whole language approaches that were seen as compatible with the NAEP framework. However, local educators were more varied in their approaches to reading, and those who favored a phonics-based method, for example, found the NAEP models less relevant. Thus, one case study respondent from a state with strong local control of education (West Virginia) reported that he had selected the "mixed" option because NAEP's impact had varied by district within his state, with the extent of impact dependent upon the local popularity of the literature-based, whole language approach to reading instruction.

To what extent has the TSA influenced state instructional or assessment practices in reading or mathematics?

Discussions about the TSA's influence at the state level must be placed within the context of national trends in curricular and assessment reform. Both our case study and survey results indicated that high percentages of state reading and mathematics programs were undergoing changes. A number of states had begun revamping their frameworks and assessments in both subject areas in the late 1980's, and high rates of change continued into the 1990's, during the period when the TSAs were underway.

Results in table 3 indicate that over three-quarters of the states responding to the survey had made or were making changes to reading curriculum frameworks since 1990, while nearly all had made or were making such changes in mathematics. In each subject area, similar percentages reported changes in instructional delivery and assessment. The somewhat greater incidence of change in mathematics may relate to the widespread acceptance of the National Council of Teachers of Mathematics (NCTM) standards. It may also be the case that more states have been focusing their reform efforts on mathematics and science because of recent policy initiatives such as the National Science Foundation's State Systemic Initiative program.

Table 3 — *Percentages of states reporting changes in instruction and/or assessment since 1990*

	Percent of States Reporting Changes for Reading	Percent of States Reporting Changes for Mathematics
Curriculum/framework	78	96
Instructional delivery	76	93
Preparation/certification of teachers	29	64
Assessment	84	91
Number Responding	45	45

Source for math: Since 1990 when the first NAEP TSA in mathematics was administered, has your state made/begun changes in any of the following? [1995 Impact Study Questionnaire]
Source for reading: Since 1992 when the first NAEP TSA in reading was administered, has your state made/begun changes in any of the following? [1995 Impact Study Questionnaire]

Fewer states reported changes in teacher preparation and certification, although here again changes were more frequent for mathematics than for reading: about two-thirds of responding states reported changes in mathematics teacher preparation, whereas less than one-third reported changes for reading.

Specific changes in reading. Within the broad areas described in table 3, certain specific types of change seemed to predominate. For reading, changes to each of the following aspects of reading instruction were reported by more than 70 percent of the states.

- More emphasis on higher-order thinking skills,
- Better alignment with current research on reading,
- Development of a standards-based curriculum,

- More emphasis on literature, and
- Better integration or alignment of assessment and instruction.

Many of these changes parallel influences in the NAEP reading assessment, and reflect "progressive" trends in reading instruction. By contrast, a much lower percentage of states reported changes in the direction of greater emphasis on phonics or basic skills (table 4).

Table 4 — *Percentages of states reporting specific types of changes in reading curriculum, instruction, or teacher preparation*

	Percent of States Checking Response
More emphasis on higher-order thinking skills, construction of meaning, and/or reader response	89
Better alignment with current research on reading	82
Development of a standards-based curriculum	78
More emphasis on literature	76
Integration/alignment of assessment and instruction	73
More emphasis on phonics and basic skills	22
More stringent requirements for teacher certification	15
Number Responding	45

Source: Which of the following characterize the ... changes [in reading instructional program (including curriculum, instruction, and teacher preparation)] that were/are being made? (Mark all that apply). [1995 Impact Study Questionnaire]

Table 5 shows that reading assessments also have been subject to dramatic changes. Greater emphasis on higher-order thinking skills, development of student performance standards, better alignment with current research on reading, and better integration or alignment of assessment and instruction were the most commonly cited changes. In general, the reading specialists who were interviewed as part of the case studies verified these changes to state reading instruction and assessment programs.

Table 5 — *Percentages of states reporting specific types of changes in reading assessment*

	Percent of States Checking Response
More emphasis on higher-order thinking skills, construction of meaning, and/or reader response	73
Development of student performance standards	73
Better alignment with current research on reading	64
Integration/alignment of assessment and instruction	64
More use of authentic passages	56
Greater inclusion of students with disabilities	53
More use of constructed response items	49
More emphasis on literature	48
More emphasis on phonics and basic skills	13
Greater inclusion of second language learners	42
Number Responding	45

Source: Which of the following characterize the changes [in state reading assessment] that were/are being made? (Mark all that apply). [1995 Impact Study Questionnaire]

Specific changes in mathematics. The NCTM standards, released in 1990, have had a profound effect upon mathematics education in the United States, and they have also strongly influenced the NAEP mathematics frameworks. Changes in mathematics instruction and assessment, summarized in tables 6 and 7, reflect this influence. Table 6 shows that alignment with the NCTM standards, greater emphasis on higher-order thinking skills or problem-solving, and development of a standards-based curriculum were the most common types of changes in mathematics instruction, reported by nearly all of the responding states. These were followed by integration or alignment of assessment and instruction, reported by 89 percent. In terms of assessment, the same four types of changes were reported by more than three-quarters of the responding states, as shown in table 7.

Table 6 — *Percentages of states reporting specific types of changes in mathematics curriculum, instruction, or teacher preparation*

	Percent of States Reporting Change
Better alignment with NCTM standards	98
More emphasis on higher-order thinking skills or problem-solving	98
Development of a standards-based curriculum	96
Integration/alignment of assessment and instruction	89
More emphasis on basic concepts and skills	36
More stringent requirements for teacher certification	27
Number Responding	45

Source: Which of the following characterize the changes that were/are being made? [1995 Impact Study Questionnaire]

Table 7 — *Percentages of states reporting specific types of changes in mathematics assessment*

	Percent of States Reporting Change
More emphasis on higher-order thinking skills or problem-solving	82
Better alignment with NCTM standards	80
Development of student performance standards	78
Integration/alignment of assessment and instruction	76
Increased use of calculators	73
Increased use of constructed-response items	64
Greater inclusion of students with disabilities	60
Increased use of hands-on activities	51
Greater inclusion of second language learners	40
More emphasis on basic concepts and skills	22
Number Responding	45

Source: Which of the following characterize the changes that were/are being made? [1995 Impact Study Questionnaire]

As these results show, *curriculum/framework, instructional delivery, and assessment practices for reading and mathematics are apparently undergoing change in most states. Most of the changes are compatible with recent directions taken by NAEP.*

Teacher preparation and certification practices, on the other hand, do not seem to be keeping pace with the rest of the changes. It may be the case that changes in certification or preparation take place outside the state department of education, and therefore state personnel (who were our informants in this study) are unaware of these developments. In our conversations with assessment directors and specialists in the case studies, some state personnel did profess ignorance when asked about teacher preparation. A number were involved with teacher in service training, however, and had conducted, or were preparing to conduct, workshops to assist teachers with new performance assessment techniques. Furthermore, case study respondents, such as those from Rhode Island, confirmed the perception that professional development efforts often cannot keep pace with changes in instructional delivery and assessment.

NAEP influences on change. How much of this change can be attributed to NAEP? As it turns out, the majority of our survey respondents felt that NAEP had an influence, albeit a minor one, on changes in their respective states.

Reading. Table 8 tallies the reported influence of the TSA program on reading instruction and assessment, for states that had participated in at least one TSA in reading and that reported any changes in reading instruction or assessment since 1990. Because as many as a quarter of the respondents professed ignorance of the extent of NAEP influence in one or all areas of reform, the responses of the curriculum and assessment respondents were aggregated, and states were classified in accordance with the highest estimate of influence indicated by either respondent.

Table 8 — ***Amount of influence the TSA program had on changes occurring in reading, among states that participated in at least one TSA in reading and that reported any changes in reading instruction or assessment***[1]

	Reading Instruction	Reading Assessment
Major influence	8	20
Minor influence	66	49
No influence	13	17
Don't know/other	13	14
Number Responding	38	35

Source: How great was the influence of the NAEP TSA on the decision to change and/or the types of changes that were made? Influence on reading instructional program, including curriculum, instruction, and teacher preparation/influence on state reading assessment program. [1995 Impact Study Questionnaire]
[1]Highest estimate for state.

Respondents from only eight percent of the eligible states credited NAEP with a major influence on changes occurring in reading instruction. An additional 66 percent, however, indicated a lesser degree of influence for NAEP. In reading assessment, the percentage of these states citing a major influence was

somewhat higher (20 percent), but the percentage citing a minor influence was correspondingly lower (49 percent), so that the overall attribution of change was about the same.

Tables 9 and 10 present more detail on the type of influence attributed to NAEP in reading. As can be seen in table 9, influence was most frequently attributed to the assessment items and framework, which served as models for the states' own efforts. Credit was also given to the general heightening of awareness caused by TSA publicity, but only about a quarter of the states reported an influence arising specifically from their own reading results in either 1992 or 1994.

Table 9 — *Aspects of the TSA program that influenced reading instruction and/or assessment, among states that participated in at least one TSA in reading and that reported any changes in reading instruction or assessment*

	Percent of States Reporting Influence
Form of the TSA assessment or types of items	46
NAEP framework	44
General heightening of awareness caused by TSA publicity	38
State's 1994 TSA reading results	26
State's 1992 TSA reading results	23
Number Responding	39

Source: Which aspect(s) of the TSA contributed to its influence on your state's reading instruction and/or assessment? *[Mark all that apply]* [1995 Impact Study Questionnaire]

Table 10 shows that the TSA program was most frequently credited with reinforcing the validity of reading changes already contemplated or underway. Nearly 70 percent of the eligible states made this attribution. This was followed in frequency by reports of the program's utility for educating local educators about planned or needed changes. About a third of the states indicated that NAEP had provided new ideas of what to change, and a similar percentage reported that the TSA helped education planners sell change to policy makers and legislators. By contrast, only eight percent reported that the TSA caused policy makers or legislators to press for changes not endorsed by education planners.

BEST COPY AVAILABLE

Table 10 — ***Specific influences of the TSA program on instruction and/or assessment in reading, among states that participated in at least one TSA in reading and that reported any changes in reading instruction or assessment***

	Percent of States Reporting Influence
Reinforced validity of changes already contemplated or underway	69
Was useful to educate local educators about planned or needed changes	49
Gave us new ideas of what to change	36
Helped education planners sell change to policy makers/legislators	31
Convinced education planners that change was needed	23
Caused policy makers/legislators to press for changes not endorsed by education planners	8
Number Responding	27

Source: How would you characterize the influence of the TSA on your state's reading instruction and/or assessment? *(Mark all that apply.)* [1995 Impact Study Questionnaire]

These survey findings were validated by the case studies. According to the assessment directors and reading specialists we interviewed for the latter, changes in assessment and/or instruction that were undertaken during the 1990's were already "in the works" by the time the NAEP reading framework was developed in 1990 and by the time the first TSA reading assessment was administered in 1992. Although the TSA program did not directly instigate these changes, it *did* serve to reinforce the validity of changes that were being contemplated at the state level, or were already underway. To a lesser extent, the TSA helped state-level educators to promote certain types of change by providing examples and external validation to support the direction of change.

Mathematics. Results for mathematics were generally similar. As can be seen by comparing table 11 with table 8, slightly higher proportions of the states credited the TSA program with influence in mathematics than in reading. However, the general pattern of responses (e.g., higher percentages of major influence reported in assessment compared to instruction) were the same.

Table 11 — *Amount of influence the TSA program had on changes occurring in mathematics, among states that participated in at least one TSA in mathematics and that reported any changes in mathematics instruction or assessment*[1]

	Mathematics Instruction	Mathematics Assessment
Major influence	13	22
Minor influence	68	57
No influence	13	11
Don't know/other	8	11
Number Responding	40	37

Source: How great was the influence of the NAEP TSA on the decision to change and/or the types of changes that were made? Influence on mathematics instructional program, including curriculum, instruction, and teacher preparation/influence on state mathematics assessment program. [1995 Impact Study Questionnaire]
[1]Highest estimate for state.

Comparison of tables 12 and 13 with 9 and 10 again shows that similar patterns of influence were reported in mathematics as in reading, although the proportions of states reporting each source or type of influence were generally slightly higher in mathematics. Notable in table 12 is the extent of influence that was attributed to the general heightening of awareness caused by TSA publicity. In keeping with the greater publicity accorded to the very first TSA, which was in mathematics, 63percent of states reported this as an influence for changes in mathematics, compared to only 38percent of states for reading.

In table 13 another dissimilarity from reading is evident. Here we see that fully half of the eligible states credited NAEP with giving them new ideas of what to change in mathematics. This increase over the percentage making a similar attribution in reading likely reflects the timing of the first TSA in mathematics, which followed rather closely on the release of the NCTM standards.

Table 12 — *Aspects of the TSA program that influenced mathematics instruction and/or assessment, among states that participated in at least one TSA in mathematics and that reported any changes in mathematics instruction or assessment*

	Percent of States Reporting Influence
General heightening of awareness caused by TSA publicity	63
NAEP framework	55
Form of the TSA assessment or types of items	55
State's 1992 TSA mathematics results	35
State's 1990 TSA mathematics results	30
Number Responding	40

Source: Which aspect(s) of the TSA contributed to its influence on your state's mathematics instruction and/or assessment? *[Mark all that apply]* [1995 Impact Study Questionnaire]

Table 13 — *Specific influences of the TSA program on instruction and/or assessment in mathematics, among states that participated in at least one TSA in mathematics and that reported any changes in mathematics instruction or assessment*

	Percent of States Reporting Influence
Reinforced validity of changes already contemplated or underway	73
Gave us new ideas of what to change	50
Was useful to educate local educators about planned or needed changes	50
Helped education planners sell change to policy makers/legislators	38
Convinced education planners that change was needed	30
Caused policy makers/legislators to press for changes not endorsed by education planners	10
Number Responding	40

Source: How would you characterize the influence of the TSA on your state's reading instruction and/or assessment? *(Mark all that apply.)* [1995 Impact Study Questionnaire]

What specific contextual factors influenced the impact of the TSA results in individual states?

This section of the report draws from the nine case studies conducted by AIR staff in December, 1995. These case studies were intended to look in greater detail at the ways in which impact was affected by each state's unique needs and circumstances.

Several themes emerged from the earlier round of case studies that were conducted by AIR staff in 1992. The first of these themes was that the *states are the primary consumers of the TSA data*. This continues to be the case. State personnel use the data for program planning, and are primarily responsible for bringing NAEP results to the attention of other state-level policy makers and to educational staff at the local levels.

A second theme concerning the impact of the 1990 TSA was that *it was greatest in states that performed worst on the 1990 TSA*. With respect to the 1994 TSA, Hawaii did report that it heightened public awareness of unsatisfactory student performance in reading and spurred reform efforts in this area. On the other hand, in North Carolina, Rhode Island, and Wisconsin, *state respondents were pleased with their TSA performance and attributed their high state averages to recent reforms in reading instruction and assessment*. In other high-performing states, such as Wyoming, state-initiated reforms in education were launched independently of any assessment results. So, with respect to reading, poor performance did not initiate reforms (because those reforms had already begun), but good performance did serve to validate reform efforts.

The media value of bad news, however, was underscored by the Panel's survey of news articles that followed the release of the *First Look* reading report in April 1995. The eleven states with the most coverage were those that had performed poorly relative to other states or relative to their own performance in 1992. Only Nebraska and Wyoming, two states that performed well, and—importantly—that had no state assessments of their own, emphasized the positive.

A new theme which has emerged from our study of the 1994 TSA is that *the impact of the reading assessment seems to have been mediated by the extent to which instruction is subject to local control*. For example, in Louisiana, Rhode Island, and West Virginia, which are strong local-control states, reading instructional delivery is reported to vary widely, with curricular emphases run the gamut from phonics to the whole language approach. In states such as these, NAEP has been influential in some districts, where it is closely aligned with classroom practice, and has had very little effect in other areas. This may account for the survey report of "mixed" impact from the reading TSA.

The decentralization of instructional choices may be expanding. North Carolina and Wyoming, for example, reported recent trends toward site-based management. In these states, staff at the state level are being encouraged to reduce their focus on monitoring educational progress and to increase their efforts to facilitate and assist local districts with their instructional and assessment programs. This seems to indicate that monitoring or accountability functions are being shifted to the local level. Other state departments of education reported that they are beginning to serve as technical assistance centers, including Louisiana and Rhode Island.

Some state departments of education also have undergone massive reorganizations and have suffered severe cuts in funding and personnel in recent years. A case in point is North Carolina, which had its staff cut by 40 percent in the most recent round of fiscal trimming. In spite of this upheaval, North Carolina continues to be a strong supporter of the NAEP program. State personnel view it as the primary vehicle for national comparisons, which are necessary for accountability purposes. Part of North Carolina's strong support for the program may be rooted in the fact that its own instructional and assessment systems, which are closely aligned with each other, are also closely wedded to NAEP. Department personnel believed that in 1992 the NAEP reading framework and assessment represented

national standards for language arts more clearly than anything else available at the time, and are still strongly committed to NAEP's vision and leadership.

Some of the states we interviewed reported that NAEP had not had a significant influence on their state's reading curriculum or assessment. In one state, Pennsylvania, the state reading framework was already well-established by 1992. State officials are currently planning to set achievement levels, revise the content standards to align with the achievement levels, and then recalibrate both the assessment and the achievement levels. After this cycle has been completed, NAEP's influence on the state reading framework may turn out to be much greater than it is at present. West Virginia officials predict a similar state of affairs after guidelines for the new reading curriculum framework are agreed upon at the upcoming legislative session. That is, they expect that NAEP's influence will *increase* with the new round of revisions to the framework.

Where NAEP has had an influence on assessment, it has affected the formats of state assessments, leading to the inclusion of more open-ended and extended response items and an increasing emphasis upon the use of authentic texts and passages. States also cite NAEP as a leader in the movement toward assessing higher-order thinking skills on reading tests. Trends toward the inclusion of greater numbers of students with disabilities and second language learners are also attributed to NAEP's leadership.

The following conclusions may be drawn from the case study interviews:

- One of the reading TSA's main roles has been to reinforce, or validate, changes in reading curriculum or instruction that were already underway in 1992 (when the reading TSA program was begun).

- States that performed well on the reading TSA view the program favorably, and report that it confirms that their own state reading programs are on the right track and are having a positive impact on student learning.

- In states with strong local control over education, the TSA's impact is uneven or difficult to evaluate. On the other hand, in states with strongly decentralized educational assessment systems, the TSA is often the only vehicle for national comparisons.

- Two states reported concerns about future participation in the TSA. In Pennsylvania, competition from the state's own assessment system, which is aligned with instruction and provides district-, school-, and student-level results, is leaching support away from the TSA. Rhode Island has similar worries about local support for the TSA. Because of its small size and population, the TSA (as currently structured) samples most of the schools in the state during any given cycle. This places a heavy burden on schools, and makes school recruitment very difficult, because schools are constantly being approached to participate in the TSA.

How highly do consumers value NAEP as a monitor of education programs?

One observation that emerged from the 1992 case study analysis was that NAEP is primarily valued for its role in sustaining and supporting broader trends in education reform. This perspective continues to be evident, with several states in the 1995 case study interviews reporting that they consider NAEP as a reference point for curricular and assessment reform in their states. For example, according to Dr. Doug Rindone, Connecticut's assessment director, NAEP provides a nationally reviewed and respected framework that "can't be ignored" in the process of developing models of curriculum, instruction, and

BEST COPY AVAILABLE

assessment for his state. NAEP continues to be viewed as a high-quality indicator of academic achievement, which makes it an invaluable tool for accountability and for national comparisons. States (such as North Carolina) cited the desire to align their state's frameworks and assessments with national standards in reading as a reason for using NAEP as a model.

As another example, although the TSA has not directly influenced curriculum, instruction, or assessment in Wyoming, it has filled a void created by the lack of a state assessment program, and provided a means for the state to measure its academic achievement over the years in relation to itself and other participating states. In addition, NAEP has always rewarded the state with a high ranking among the participating states, so Wyoming have a very positive attitude toward the role of NAEP in their state.

What are perceived to be the TSA's major weaknesses?

Despite the positive values associated with NAEP, more than half of the survey respondents also pointed out problems that limited its utility. For reading, 64 percent of responding assessment directors and 53 percent of responding reading specialists from states that had participated in at least one reading TSA felt that there were specific problems with the NAEP TSA assessments that limited their utility in the states; for mathematics the corresponding percentages were 55 percent and 53 percent (table 14).

Table 14 — *Percent of respondents reporting that problems with the TSA limited its utility to the states, among states that participated in at least one TSA in the relevant subject*

	TSA in reading		TSA in mathematics	
	Percent of Assessment Directors	Percent of Reading Specialists	Percent of Assessment Directors	Percent of Mathematics Specialists
Problems	64	53	55	53
No problems	21	25	35	28
Don't know/no response	15	22	10	19
N	39	36	40	36

Source: Do you feel that there have been specific problems with NAEP TSA assessments in reading/mathematics that have limited their utility for your state? [1995 Impact Study Questionnaire]

The specific types of problems identified are tabulated in tables 15 and 16.

Table 15 — *Specific problems with the reading TSA reported as having limited its utility to the states, among states that participated in at least one TSA in reading*

	Percent of Assessment Directors		Percent of Reading Specialists	
	Major Problem	Minor Problem	Major Problem	Minor Problem

Not sufficiently aligned with current research	3	3	0	6
Not sufficiently aligned with state curriculum	5	3	0	11
Not sufficiently aligned with classroom practice	5	5	8	6
Not sufficiently helpful for diagnosing instructional problems	23	8	14	3
Too much lag time to reporting	44	13	25	6
Provides no local or district results	41	13	17	8
Assessment schedule unpredictable	28	10	14	8
Between-state comparisons do not control for demographics	15	15	11	6
N	39	39	36	36

Source: Do you feel that there have been specific problems with NAEP TSA assessments in reading that have limited their utility for your state? If there have been problems, mark all that apply: [1995 Impact Study Questionnaire]

Table 16 — *Specific problems with the mathematics TSA reported as having limited its utility to the states, among states that participated in at least one TSA in mathematics*

	Percent of Assessment Directors		Percent of Mathematics Specialists	
	Major Problem	Minor Problem	Major Problem	Minor Problem
Not sufficiently aligned with current research	3	0	0	3
Not sufficiently aligned with state curriculum	0	3	6	3
Not sufficiently aligned with classroom practice	0	10	3	3
Not sufficiently helpful for diagnosing instructional problems	20	8	11	3
Too much lag time to reporting	40	8	25	8
Provides no local or district results	40	10	25	17
Assessment schedule unpredictable	33	5	8	8
Between-state comparisons do not control for demographics	13	15	11	11
N	40	40	36	36

Source: Do you feel that there have been specific problems with NAEP TSA assessments in mathematics that have limited their utility for your state? If there have been problems, mark all that apply: [1995 Impact Study Questionnaire]

For both assessments and each of the respondent groups, lag time to reporting and lack of local district results stand out as the major sources of problems. Assessment directors also expressed concern regarding the unpredictability of the TSA assessment schedules and, to a lesser extent, with the fact that the assessments are not sufficiently helpful in diagnosing specific instructional problems. By contrast, very few of the respondents faulted the assessments for their overall designs or frameworks, as would have been the case if they had cited lack of alignment with current research, NCTM standards, state curricula, or classroom practice as problems.

BEST COPY AVAILABLE

Case study results. Case study results once again allowed us to elaborate on the findings from the survey. Case study states were nearly unanimous in their opinions that the lag time in reporting TSA results is too long. An exception to this rule was offered by the state of Wisconsin's reading specialist, who felt the lengthy lag time sends a message to consumers that reporting high-quality results from a complex assessment program requires careful analysis and takes a substantial amount of time.

Several case study respondents felt that it is particularly difficult to justify participation to districts and schools in the absence of results that relate directly to them. Others mentioned that the TSA is not useful for individual or district diagnostic purposes. Consequently local staff, as well as students, perceive TSA administration as a burden. Claudia Davis (Louisiana) pointed out that districts were being asked to participate in the 1996 TSA before state data from the 1994 reading assessment were made available (only the *First Look* reports had been issued by recruitment time, not the state reports).

Other limitations on the TSA that were cited by one or more case study respondents included the following:

- The *unpredictability of the assessment schedule* was seen as a problem not only in limiting the TSA's use but also for recruiting schools. Locals would be able to better plan for their district and school testing programs if they knew which areas and which grades were to be tested far in advance of the administration date. NAEP results would be able to play a more major role in planning if districts and schools could be assured that the needed information would be available. The unpredictable assessment schedule has also made it more difficult for states to recruit schools. If schools don't participate, NAEP loses most, if not all, influence in the district. In order to get cooperation of districts and develop interest in the results, the state must be able to communicate to districts and schools on an ongoing basis regarding the assessment results (grade levels and content areas) locals can expect to have for their use.

- *In areas where classroom practices do not reflect NAEP, the TSA has limited usefulness.* This is especially true in states where classroom practice varies widely by district, as is the case in strong local control states.

- A *better description of the performance standards and sample items* should be shared with the public when the results are published. It is not clear what the levels, such as proficient or advanced, mean, and this lack of understanding limits the usefulness of the TSA data.

- Another factor limiting NAEP's utility as a model assessment at the local level is that, because of NAEP's sampling design, *only those teachers who actually participate in its administration have the opportunity to explore the TSA thoroughly.*

- *The between-state comparisons make limited allowance for factors beyond educators' control, thus limiting the utility of the data.* Some states feel that NAEP reporting should explicitly control for the possible effects of demographic factors so that the influence of these contextual factors (and the validity of the TSA results) will not be questioned.

In addition to the limiting factors enumerated above, factors beyond NAEP's control often impact its usefulness at the state level. For example, some parent groups are opposed to the background questions asked on the TSA, particularly questions about family and the amount of television children watch. Therefore, Pennsylvania, a state where such groups are particularly active, does not require students to answer the NAEP background questions. Although this fix makes it possible for Pennsylvania to continue to participate in the TSA program, it does limit the validity and reliability of the background data that are collected for the assessment.

Case study participants in Wisconsin pointed out that the utility of the information collected through NAEP has been limited by a number of factors which are beyond the control of the assessment. In some cases, NAEP is being asked to be things that it was never intended to do, such as to diagnose instructional problems. In other cases, competing priorities may hinder efforts to address limiting factors; for example, decreasing the lag time for reporting results may substantially affect the amount and quality of the results reported. Furthermore, the NAEP TSA collects a tremendous amount of data that could be useful to states but states lack the capacity to use these data; furthermore, they lack the resources to disseminate them to the schools and to the public. The state assessment director in Wisconsin, Dr. Darwin Kaufman, would like to see how states can work with those at the federal level to figure out how to report information in ways that will make a difference and have people take notice.

State testing directors survey. In the 1994 State Testing Directors survey, Panel staff asked respondents in both participating and nonparticipating states to react to the following question: *What is the principal threat to the success of the State NAEP and how could this problem be addressed?* Problems cited by the participating states fell into the following categories: cost, securing school participation, the lack of district data, NAEP's lengthy analysis-and-reporting schedule, the TSA's unpredictable assessment time line, competition from a state's own assessment program, and political imbroglios, such as conservative opposition to government testing programs. States that had not participated in the 1994 TSA cited the lack of financial resources, competition from their own assessment programs, lack of feedback on schools, and negative publicity from previous TSA participation as major barriers.

Has participation in NAEP been viewed as a worthwhile exercise? Do states plan to participate in the future?

Table 17 shows that three-quarters of the impact survey respondents thought that future participation in the TSA program would be at least somewhat worthwhile (76 percent of assessment directors, 80 percent of reading specialists, and 75 percent of mathematics specialists). Interestingly, reading specialists were the most likely to rate future participation as at least somewhat worthwhile, but least likely to rate it as very worthwhile. In the State Testing Directors survey mentioned above, 92 percent of respondents from states that participated in the 1994 TSA indicated that they planned to sign up for the 1996 State NAEP, some pending budgetary approval. Furthermore, half of the respondents from nonparticipating states indicated a willingness to participate in 1996. *These levels of endorsement seem to indicate a general sense of satisfaction with the program, especially among current participants.*

Table 17 — *Evaluation of the value of participation in future state NAEP assessments, among states that participated in at least one TSA*

	Percent of Assessment Directors	Percent of Reading Specialists[1]	Percent of Math Specialists[2]
Very worthwhile	38	19	42
Somewhat worthwhile	38	61	33
Not worthwhile	13	3	8
Don't know	10	17	17
N	39	36	36

Source: Assuming that conditions for participating in state NAEP remain essentially as they are now, to what extent do you think that future participation for your state would be worth the time, effort, and money involved? [1995 Impact Study Questionnaire]

[1]Restricted to reading specialists from states that participated in at least one TSA in reading.

[2]Restricted to mathematics specialists from states that participated in at least one TSA in mathematics.

Furthermore, opinions of the perceived value of the TSA appear to be holding steady or rising since 1990. As shown in table 18, 50 percent of mathematics specialists and 43 percent of assessment directors and reading specialists in participating states. reported that their opinions of the TSA's overall value had become more positive. By contrast, only 10 percent of assessment directors and 3 percent of curriculum specialists indicated that their perceptions were becoming more negative. Relatedly, the implementation of the 1994 TSA received very high marks from the assessment directors responding to the 1994 State Testing Directors Survey; an overwhelming majority (92 percent) gave the assessment a grade of 'A' or 'B.'

Table 18 — *Changes in perceived value of TSA since 1990, among states that participated in at least one TSA*

	Percent of Assessment Directors	Percent of Reading Specialists[1]	Percent of Math Specialists[2]
Become more positive	43	43	50
Become more negative	10	3	3
Remained unchanged	48	51	47
N	39	35	36

Source: Since 1990 when the NAEP TSA assessments first began, has your opinion of the TSA's overall value become more positive, more negative, or remained unchanged? [1995 Impact Study Questionnaire]

[1]Restricted to reading specialists from states that participated in at least one TSA in reading.

[2]Restricted to mathematics specialists from states that participated in at least one TSA in mathematics.

Most respondents in the case studies indicated that, overall, the NAEP TSA has had a positive impact on education in their states. For example, assessment director Claudia Davis (Louisiana) believes that the TSA has grown in value since its beginning in 1990, and that it has a particularly positive impact on teachers. From the viewpoint of reading curriculum, the TSA has been found to be very worthwhile in Louisiana, but its impact is difficult to measure because of strong local district control. In West Virginia, assessment staff are looking more closely at NAEP than ever before, primarily because they want students to have hands-on experience in mathematics and in science, and NAEP provides a good model for this type of instruction and assessment.

In a few of the states, NAEP's influence has been mixed. In Pennsylvania, NAEP has had difficulty competing with the state and local assessments. Participation, in the short run, looks shaky. And although the NAEP TSA has had a "generally positive" impact to date on education in Rhode Island, which has included NAEP in its 5-year assessment plan, Rhode Island's participation in future NAEP TSAs is questionable. Because of its size and the requirement to provide a sample of 2,500 students, the burden on schools has been greater in Rhode Island than in most other states. The problem of overburdening schools, coupled with the development of performance assessments specifically for the Rhode Island assessment, and participation in efforts such as New Standards Project has increased the uncertainty about future participation in NAEP.

♦ Summary and Conclusions

Several of the conclusions that will be presented here replicate, and therefore serve to reinforce, conclusions that have appeared in earlier Panel reports on the impact of the Trial State Assessment program. Overall, it appears that the state NAEP continues to sustain its perceived value to its main constituents, state-level educators and policy makers. The TSA's overall impact on education was judged as generally positive by survey respondents, and individuals interviewed for the case studies echoed these sentiments. Furthermore, early fears about possible negative impacts from state NAEP appear to have been allayed, and sentiments about the TSA have grown more positive over the course of the program, as states have become more familiar with it and have had more experience with it.

Results of the December 1995 survey, which encompassed the perceived impacts of all three TSAs (1990, 1992, and 1994), confirmed preliminary conclusions that the Panel had drawn on the basis of the impact of the 1990 TSA. That is, the analyses once again suggested that the TSA influenced education at the state level, not in isolation, but because it articulated well with other influences in mathematics and reading education. For example, the format of the NAEP reading assessment, in particular its inclusion of extended-response items and authentic reading passages, was used to justify similar modifications in state assessment programs all over the United States. Furthermore, the justification for giving greater weight to assessing higher-order thinking skills in reading and language arts in state assessments was provided when NAEP went in this direction. NAEP continues to be viewed as a leader in assessment by the states, who are willing to follow its lead because it mirrors the best and most progressive thinking in the reading (and mathematics) communities.

With respect to specific contextual factors that determined the influence of NAEP, we have seen that the TSAs in mathematics coincided with other strategic events, such as the release of the NCTM standards, that facilitated change. In these circumstances, poor performance on the TSA added an extra spur to reform efforts at the state level. In reading, on the other hand, reform efforts in many states were further along by the time the TSAs were administered. Many state personnel did, however, stress that the NAEP was valuable for reinforcing the need for, and value of, reading reforms that were already underway in their states.

From the states' perspective, major areas of weakness in the TSA include the lengthy analysis-and-reporting schedule and the lack of district- or school-level results. With respect to the issue of impact, this suggests that the TSA might have measurably greater impact at the state level if reports were issued on a faster time line, and if those reports were relevant at the local (school or district) level. Neither of these are simple matters, however. The provision of district- or school-level results, in particular, would imply major changes in the state NAEP design and even in the program's mission.

In evaluating impact, one must bear in mind the fundamental purposes of the NAEP program, and the kinds of impact that are consonant with those purposes. NAEP is not attempting to replicate the functions of a state assessment or to produce the kinds of impact associated with high stakes testing. Reassuringly, most states indicated that, despite these perceived drawbacks, they were valuing NAEP on its own terms. They judged participation in the TSA to be well worth the time, effort, and expense, and they placed a high value on NAEP as a source of external validation for their own assessment programs, as a vehicle for enabling national comparisons, and as a model of contemporary assessment practices.

Appendix

1994 National Assessment of Educational Progress Trial State Assessment Impact Case Studies

Case Study for Connecticut

Overview

Connecticut continues to be in the forefront of change and reform in curriculum, instruction, and assessment. Since the late 1980s, with the impetus of legislative mandate, Connecticut has developed and refined a comprehensive state assessment program, incorporating current research on assessment and instructional practice. Guided by the new assessment program, the state is currently implementing a comprehensive reform of curriculum as well. Along with other models, the NAEP framework played an important role in developing Connecticut's assessment program, and it has been useful in informing local educators about planned and needed changes in curriculum and instruction. In addition, publicity about the state assessment program has created a generally heightened public awareness about current developments **in curriculum, instruction, and assessment in the state.**

The State of Education in Connecticut

Organization

Connecticut's state department of education (DOE) has primary authority for designing and implementing education policy and practice in the state. The state legislature consults with the DOE when enacting education legislation; and, according to Dr. Douglas Rindone, chief of the Bureau of Evaluation and Student Assessment, it provides "full authority" to the DOE for implementation. For example, state law requires only that basic testing be done at grades 4, 6, 8, and 10. In a collaborative process with local educators, DOE staff develop, implement, and attempt to align curriculum, instruction, and assessment for 165 LEAs across the state.

Within the Bureau of Curriculum and Instructional programs, 20 professional staff are responsible for curriculum programs in core areas, categorical programs, and special grants. Currently one language arts specialist, Ms. Angela Rose, is responsible for statewide professional development programs; development of curriculum resources, including K-12 content and performance standards in language arts; content oversight of the statewide language arts testing program; technical assistance to school districts; and working on pre-service training with post-secondary institutions. Within Dr. Rindone's 30-person Bureau of Evaluation and Student Assessment, nine professionals are responsible for all phases of developing and implementing the state assessment program (including relevant in-service training), in close collaboration with state curriculum/instruction staff, teachers, committees, and outside contractors.

Reading Curriculum and Instruction

Connecticut's reading framework is currently undergoing a major revision, its first update since 1981. With a 1996 publishing date expected, the new framework will include a statement of philosophy, a new common core of learning, a broad set of goals, and a set of performance standards that reflect student outcomes to be achieved by the end of grade 12. In addition to the need to improve an "outdated" curriculum, revision of the framework was prompted by implementation and refinement of the mandated state assessment program (described below) and the desire to achieve

BEST COPY AVAILABLE

better alignment between assessment and curriculum/instruction in the state. The new framework, which is not mandated, is intended to be used by LEAs as guidelines and recommendations for curriculum, instruction, professional development, and related issues.

With regard to public support for the new framework, Angela Rose indicated a continuing need to balance new approaches to instruction and assessment (e.g., higher-level thinking skills, literature-based learning) with adequate attention to basics (e.g., spelling, phonics), and to educate the public about the importance of creating a program based on high standards.

Assessment

Tailored to state education goals, Connecticut's assessment program includes two primary components: (a) the Connecticut Mastery Test (CMT), an assessment of reading, writing, language arts (i.e., grammar and editing), listening, and mathematics skills, which is administered to all students in grades 4, 6, and 8 in the fall each year; and (b) the Connecticut Academic Performance Test (CAPT), an assessment in mathematics, science, and language arts (i.e., response to authentic passages of literature), which is administered to all students in grade 10 in the spring each year. The CAPT also includes an interdisciplinary task, which measures how well students can read and write about current issues that have social, mathematical and scientific relevance. Prior to the science test, students are required to conduct and write up a science lab, and then respond to specific questions about the lab in an "on-demand" test.

The DOE first administered the CMT in 1985-86; administered a second-generation CMT in 1993; and will soon begin development of a 3rd-generation assessment, targeted for release in 1999. The CAPT was just administered in 1995. Both assessments are criterion- referenced and developed by the state DOE in collaboration with local educators. The reading component of the CMT also includes the Degrees of Reading Power assessment (published by Touchstone Applied Science Associates in New York, and formerly marketed by the College Board). This assessment uses the "CLOZE" technique of measuring reading proficiency, including multiple-choice items ranging from very easy to very difficult.

Unlike the three-tiered achievement levels used by NAEP, the CMT and CAPT use only one level of mastery, or goal standard, for each subject area. Standards Setting Committees, comprised of state and local representatives, use a modified Angoff technique to determine a single mastery standard for each subject and grade level.

As in many other states, Connecticut's education climate is increasingly influenced by the demand for accountability at all levels; the state assessment is a key tool for accountability. Aggregate results for the state and school districts are widely publicized by the press and media and are used to identify program weaknesses and to guide program improvement. In addition, parents and teachers receive individual student reports that further identify achievement by subskill area and are used for diagnosis and remediation at the individual student and school levels.

In addition, the 10th-grade CAPT is designed to identify students who achieve the state goals and to award these students with a Certificate of Mastery and transcript certification that they have performed with distinction in specific subject areas. According to Dr. Rindone, the CAPT is a "tough test" for which only 30 percent to 35 percent of all 10th graders receive certification— unlike graduation tests that are typically designed for most students to pass.

Influence of the NAEP TSA

Participation in NAEP is mandated in Connecticut; state schools participated in the 1990 TSA of 8th-grade mathematics and the 1992 and 1994 TSAs of 4th-grade reading. DOE staff who were interviewed for this study (Dr. Doug Rindone and Ms. Angela Rose) believe that the TSA has influenced the development of Connecticut's reading curricula and assessments in an important way. According to Dr. Rindone, NAEP provides a nationally reviewed and respected framework that "can't be ignored" in the process of developing curriculum/instruction and assessment in

Connecticut. Consequently, NAEP has served as a reference point for developments in curriculum and assessment in the state.

According to language arts specialist Angela Rose, the NAEP reading framework provided a model for (a) the CAPT language arts component, in which a student's response to authentic literature is holistically scored; (b) the CAPT interdisciplinary task; and (c) the second generation CMT for grades 4, 6, and 8. Preliminary discussions about the 3rd- generation CMT have been strongly influenced by the NAEP literacy framework. Two particularly useful features of the NAEP framework, she said, were aspects of the literacy grid and the dynamic interaction graphic (Michigan Theory).

Ms. Rose believes that NAEP's influence on curriculum and instruction has been more subtle than its influence on assessment. Unlike assessment, changes in curriculum and instructional strategies have not been mandated; they have also been gradual and less visible than changes in the content and format of the assessments. Nonetheless, the NAEP framework has impressed curriculum staff, particularly its literature aspects (e.g., construction of meaning from literature, organized into four levels, and the assessment of literature experiences).

Since Connecticut students consistently score high on NAEP relative to students in other states, TSA scores do not have as much impact in Connecticut as the TSA framework, format (i.e., types of items), and content, as described above. According to Dr. Rindone, the TSA results have been somewhat useful, primarily in reassuring the public that their students continue to do well and also in reinforcing public information about the CMT and CAPT assessments. The public, however, focuses more on results of the Connecticut state assessments.

Limitations of the TSA

Primary limitations in the TSA's utility, cited by Ms. Rose, include lag time in reporting results and lack of district-level results. Because local-level information is not reported, the TSA is not useful for individual or district diagnostic purposes. Consequently, local staff, as well as students, perceive TSA administration as a burden. Ms. Rose also indicated that the format and content of the TSA are somewhat sensitive politically in that many stakeholders do not understand performance assessment. Although Dr. Rindone did not identify specific problems with the reading TSA, he emphasized the importance of criterion-referenced testing and reporting of NAEP results. He expressed the hope that NAEP would soon resolve issues related to the developmental status of the achievement levels.

Case Study for Hawaii

Overview

Several forces have been driving changes in the "state of education" in Hawaii—notably, a legislative mandate of statewide performance standards and educators' increasing support for new restructuring curriculum and assessment practices. As a result, in Hawaii, the NAEP TSA has reinforced changes that were already underway in curriculum, instruction, and assessment. In addition, the NAEP results have increased public awareness that students in Hawaii were not performing satisfactorily in reading, and have underscored the need to focus on literacy, identify student outcomes that give direction and focus for classroom instruction, and develop assessments that align with the curriculum and provide rich information about student performance.

The State of Education in Hawaii

Organization

The Hawaii Department of Education serves 186,581 public school students (122,596 at the elementary level, and 63,985 at the secondary level). An elected state board of education formulates policy and exercises control over the public school system through its appointed superintendent of education. The public schools are organized under seven geographic district offices and managed through district superintendents.

Four major staff offices headed by assistant superintendents provide statewide professional and technical support services and programs to the public schools. The Office of Instructional Services provides curriculum and instructional support services and programs.

The development and administration of the statewide testing program is the responsibility of the Test Development Section of the Planning and Evaluation Branch. The Test Development Section coordinates the DOE's participation in NAEP.

The DOE is currently undergoing reorganization. The 1994 state legislature passed the Omnibus Education Bill that mandates restructuring and downsizing state and district offices.

Recent forces for change. According to Dr. Selvin Chin-Chance, director of the Office of Testing, and Ms. Leila Naka, language arts specialist, several initiatives have guided curriculum and assessment restructuring in Hawaii:

1. In 1989, a Task Force on Restructuring the Curriculum was convened to recommend possible changes to the essential requirements of schools. The recommendations of the Task Force resulted in additional Foundation Program Objectives and Essential Competencies, increased mathematics and science requirements, and the development of content area frameworks and curriculum guides.

2. The Hawaii Goals for Education, developed in 1990 through statewide education summits, resulted in the development of eight education goals designed to "ensure education opportunity and excellence" for all students.

3. School/Community-Based Management gave schools greater autonomy to make decisions and freed them from constraining regulations. It was an empowering process, as well as a means to decentralize school governance.

4. In 1991, the Action Plan for Improving Mathematics, Science, Language Arts, and Social Studies addressed the need to improve unsatisfactory standardized achievement test results (including those for the NAEP TSA). The Action Plan helped to rally the DOE into making a concerted effort to address the need to improve student learning.

5. Legislative and public pressure for accountability resulted in the formation of a legislatively mandated Commission on performance standards in 1991. The purpose of the commission was to determine content and performance standards and assessment measures for the state that would result in greater accountability for teaching and learning. The performance standards were formally adopted by the state board of education in 1994.

6. In 1994, the new superintendent of education identified student literacy as the focal point within the DOE and launched the Success Compact. The Success Compact, which is to be used consistently across all grade levels, is a systematic process of teaching based on how successful learners learn. The Success Compact is currently providing staff development in more than 80 schools in the state.

Reading curriculum and instruction

The *Language Arts Program Guide* was revised in 1988, and serves to guide schools in the development of their own language arts curriculum to meet the needs of their students. The guide identifies benchmark student outcomes across grades K-12 in reading and suggests curricular and instruction elements to support attainment of those benchmarks. The view of reading described in the guide represents a conceptual shift. Reading is described as interactive, constructive, and strategic. Teaching emphasizes meaning-making, conscious connections between prior knowledge and new information, "real" purposes and materials, and the use of a variety of strategies within the reading process.

Subsequent documents—the *Essential Content and Student Outcomes for the Foundation Program*—have been published to provide schools with direction and focus for classroom instruction, curriculum, and assessment.

Assessment

Hawaii's statewide assessment program currently includes the following components: (a) the Stanford Achievement Test (SAT, Th version—basic reading, mathematics, and language arts subtests), annually administered to all students in grades 3, 6, 8, and 10; (b) the Hawaii State Test of Essential Competencies (HSTEC), which is required of all students who wish to earn a high school diploma; (c) a credit-by-examination program, administered to students in grades 8-12 to earn credit in selected subjects such as foreign language, algebra, and keyboarding; and (d) the state-mandated administration of the NAEP.

In addition to these "on-demand" examinations, Hawaii has piloted and is preparing to implement across grades 4, 7, and 11, an innovative, locally developed Hawaii Writing Assessment. Until three years ago, Hawaii administered the standardized Stanford Writing Assessment at grades 3, 6, 8, and 10. The assessment process includes four phases—collection of student work over approximately seven months, selection of a fiction or nonfiction piece for revision, revision of the selected piece of writing within a standardized four to seven hour window of time, and submission of the writing for scoring using locally developed rubrics.

In collaboration with the University of California at Los Angeles and the Center for Research on Evaluation, Standards, and Student Testing (CRESST), the state has developed a performance assessment in social studies and history for students in grades 4, 5, 7, and 11. DOE curriculum and assessment staff worked collaboratively with CRESST to modify the multiple-choice format of the previous CRESST test and to create an assessment that requires students to read and analyze authentic documents and to incorporate previous knowledge and information from reading into an essay.

Influence of the NAEP TSA

Schools in Hawaii participated in the 1990 NAEP TSA of 8th-grade mathematics and the 1992 and 1994 TSAs of 4th-grade reading. DOE staff who were interviewed for this study (Dr. Chin-Chance and Ms. Naka) believe that the TSA has reinforced changes that were already in process in Hawaii, such as the development of content and performance

standards; the increased emphasis on literacy and reading; and greater use of alternative assessments. Specifically, they believe that the 1992 NAEP results (and preliminary, slightly improved results for the 1994 reading TSA) have heightened consumers' and educators' awareness of unsatisfactory student performance and the continuing need to focus on literacy and student learning and to improve students' outcomes. They believe that the NAEP is a good alternative to current norm-referenced multiple-choice tests—particularly, the use of more open-ended items, extended reading passages, and written performance tasks.

Limitations of the TSA

Dr. Chin-Chance and Ms. Naka noted two primary problems that have limited the utility of NAEP in Hawaii: (a) the absence of local-level reporting needed for distinguishing regional differences and for assessing areas of strength and weakness, and (b) the significant lag time in reporting TSA results. Hawaii's schools serve a broad range of students with diverse characteristics. Neither the statewide NAEP results nor between-state comparisons capture these differences. As a result, Dr. Chin-Chance believes that NAEP scores are vulnerable to criticism. He also noted that the lag time in reporting NAEP results does not allow Hawaii to enact changes prior to the next NAEP administration.

Another factor limiting NAEP's utility is its sampling design. Only those teachers who actually participate in its administration have the opportunity to explore its usefulness for improving teaching and learning. Exposing more teachers to the test format, perhaps by providing sample booklets to all teachers in the state, would enhance its more widespread utility.

Despite these perceived limitations, DOE staff appear to understand the logistical barriers involved and believe that, overall, the NAEP TSA has been congruent with other state initiatives in curriculum and assessment and will have a positive impact on education in Hawaii.

Case Study for Louisiana

Overview

Public education in Louisiana is increasingly determined by the local districts (parishes). Information from the state, including scores from the NAEP TSA, is passed on to local parishes for review. Because of the emphasis on local decision making and the variation it accommodates, the impact of the TSA is difficult to estimate but considered positive overall.

State of Education in Louisiana

According to Claudia Davis, section administrator in student assessment, the role of the state's department of education is becoming less focused on monitoring and more focused on facilitating and assisting local districts. The state wants local districts to be informed about their students and to have the flexibility to make the best decisions for that student population, she says. The state has curriculum standards in place; however, a standards development Task Force is developing new standards for the core content areas (language arts, math, social studies, science, the arts, and foreign language). The local districts will decide how to implement the new standards through local curriculum development.

The state is also being influenced by national trends, like inclusion and the national goals specified in the Goals 2000 initiative. Louisiana wants all of their students to have the same opportunities for learning and to meet challenging standards. Through the coordinated efforts of the Louisiana Department of Education (LDE) and the Louisiana Systemic Initiative Program (LaSIP), a five-year program ending in 1995-96, the state has focused on improving mathematics and science instruction in grades K-12.

Organization

Cutbacks over the years have reduced staff levels in the LDE. For curriculum, the department has the Bureau of Elementary Education (grades K-8) and the Bureau of Secondary Education (grades 9-12). Also included is Starting Points, a federally funded program for children from four-year-old through kindergarten age. The curriculum staff provide technical assistance, such as workshops in whole language, upon request by local school districts. One of the current efforts is a Primary School Initiative, which is currently looking at developmentally appropriate practices across the curriculum in terms of multi-age grouping, multi-ability grouping, and assessment in the early primary grades (K-3). The Primary School Initiative has also focused on peer relationships and parental involvement. The Bureau of Pupil Accountability oversees all state-mandated testing, kindergarten screening, and participation in the NAEP TSA.

When curriculum and assessment must be aligned, for example in the ongoing standards development, the two staff groups coordinate their work closely. Louisiana has a tradition of this type of coordination, and Claudia Davis says that the coalition-building is expanding. For the standards development, coordination is occurring not only between the Assessment Department and the Bureaus of Elementary and Secondary Education, but with staff from all of the offices within the LDE.

Curriculum and Instruction

Louisiana is moving away from a state-mandated curriculum and moving toward state-defined standards and benchmarks. The state has curriculum standards that were developed collaboratively with educators from all over the state. The standards in the core content areas (language arts, math, social studies, science, the arts, and foreign language) are currently being rewritten. Math and science standards are nearly complete; those in the other core content areas should be complete by 1997.

BEST COPY AVAILABLE

Local school districts are required to meet the standards but have discretion as to how this is done. The state provides curriculum guides, including recommended activities for implementing content standards, and local school districts select the methodology and methods they will use to provide instruction. The state also adopts textbooks, and local school districts choose their textbooks from the list.

As a result of an Eisenhower grant received in 1992, Louisiana's mathematics standards are now in line with those of the NCTM, with more emphasis on complex reasoning and problem solving as well as basic concepts. The teacher preparation and certification requirements have been increased, and the Louisiana Systemic Initiative Program has developed new strategies for teachers to use in providing mathematics instruction. Model lessons called MICAS—models for integrating curriculum and assessment which are based on the math framework—have been developed by the state through a private contractor; the MICAS as well as other model lessons have been disseminated across the state for use in classrooms.

In reading, changes in the curriculum frameworks will result largely from Title 1 and Goals 2000 funds and from the realignment of funds. Susan Johnson, section administrator in the Bureau of Elementary Education, points out that reading instruction is determined by local school districts, so that changes occurring in reading instruction are not known at the state level. A state law mandates structured phonics-based reading programs for dyslexic children. Methodologies, such as whole language, are sometimes controversial at the local level.

Assessment

The Louisiana Education Assessment Program (LEAP) was implemented in 1989 and includes both norm- and criterion-referenced tests. The California Achievement Test, fifth edition, is administered to students in grades 4 and 6. State criterion-referenced tests (CRTs), based on the state curriculum guides in language arts and mathematics, are administered to students in grades 3, 5, and 7. Students in grade 7 also take a written composition test when funds are available. A bank of items was developed for the CRT in 1989, and a percentage of the test is revised every year. A Graduation Exit Exam (GEE) in English/language arts, math, and written composition is administered at grade 10; the science and social studies components of the GEE are administered at grade 11. Students have several opportunities to re-take sections that they did not pass.

All of these assessments are used primarily for diagnostic purposes, providing schools with data to assist them in planning instruction and making promotion and retention decisions. The GEE is a high-stakes test: to graduate from high school, all Louisiana public school students must pass all five of its parts, in addition to meeting the required Carnegie units. A number of local education agencies use the data from the LEAP CRTs for professional evaluation also, but this is a local decision, not a state policy. Private schools have the option to offer the GEE, with oversight by the state department of education, and about 30 private schools do so.

Mathematics assessment has not changed since 1989 because new standards have just been developed and are not yet approved by the state board of elementary and secondary education. Work on the math assessment framework is scheduled to begin in 1996. The state assessment group hopes to incorporate increased use of constructed-response items, but it is concerned about the potential expense involved. They are encouraging the use of constructed-response items at the local level. Calculators are not allowed in assessments because the state cannot ensure that all students will have access to them. An exception is made for students whose individualized education plans specify calculator use on the assessment.

The state does not report according to levels of proficiency at the present time, but is moving in that direction, Claudia Davis reports. The performance standards are reported in terms of scaled scores, which are essentially pass/fail scores. Reporting methodology will change after new standards have been developed in all of the core content areas. By the year 2000, the state hopes to have new assessments in place to reflect the new content standards.

The Influence of NAEP on Curriculum, Instruction, and Assessment

The impact of the NAEP TSA is difficult to assess because of the central role of local education agencies in Louisiana. The TSA reports are widely disseminated, but no data are available on their use. Staff members in the state department of education find the reports useful, particularly for presentation visuals, but some local instructional staff have said that they find them to be voluminous and not user-friendly.

The NAEP TSA has had a definite impact on Louisiana's state mathematics assessment. Claudia Davis reported that Louisiana piloted a math assessment at grades 5 and 7. The committee and assessment group that worked on developing this assessment really liked the NAEP math framework and assessment, and their pilot test instrument consisted primarily of released items from NAEP. The NAEP mathematics framework and item types reflect the change needed in the existing mathematics assessments. As a result of the extremely long turnaround time for NAEP results in reading, effects on assessment in reading are inconclusive at this time.

Limitations of the TSA

The primary concern expressed about the NAEP TSA is the long turnaround time in getting reports of results. The state department of education is placed in a bad position when they ask local districts and teachers to take instructional time for something that does not give them results. Claudia Davis points out that districts were being asked to participate in the 1996 TSA before state data from the 1994 reading assessment were available. Only the *First Look* report had been issued, not state reports.

Overall Evaluation of the TSA Program

Participation in the NAEP TSA is mandated by the state. The TSA has not created any specific problems in Louisiana, but Claudia Davis is concerned that the long turnaround time will affect schools' willingness to participate, particularly among private schools. While she believes that the impact of the TSA on education in Louisiana has been limited, she believes that the TSA has grown in value since its beginning in 1990. She finds that it has a particularly positive impact on teachers who have administered the TSA because some items reflected different approaches to instruction. From the viewpoint of reading curriculum, Susan Johnson also finds that the TSA is very worthwhile in Louisiana but also noted that impact is difficult to measure because of strong local district control.

Case Study for North Carolina

Overview

As in many other states, North Carolina's state department of education has undergone massive cuts, changes, and reorganizations in the past five years. In spite of this upheaval, North Carolina continues to be a strong supporter of the NAEP TSA. In an assessment climate increasingly influenced by the need for accountability, as well as trends toward site-based management and local control, NAEP continues to be the primary vehicle for national comparisons. NAEP also continues to exemplify national standards for North Carolina, where state content-area frameworks and assessment systems are very similar to NAEP's.

The State of Education in North Carolina

Organization

North Carolina's department of education is governed by the state board of education, which has primary responsibility for making education policy decisions in the state. North Carolina's legislature has historically played an active role determining what students should know and be able to do and has spearheaded demands for a high level of accountability for public schools. The state board, on the other hand, decides what *kinds of* accountability programs the schools will have, and at what level (e.g., district versus school) accountability resides.

In the past five years, the state's department of education has undergone three major reorganizations. Under the most recent legislatively mandated reorganization, begun in June, 1995, the department was cut by 40 percent. The staff of 800 was reduced to 475, and the Accountability division (wherein the assessment staff reside) was moved into the area of Instruction and Accountability Services, which has three other subdivisions: Curriculum and Instruction, School Improvement, and Exceptional Children. Accountability has a staff of 29, and Curriculum has a staff of about 50.

The North Carolina Department of Education has a regional structure, although this structure will change as part of the current reorganization. Presently, each region of the state has an assessment coordinator from the state Accountability division, and each school system has a test coordinator. The state coordinators train the test coordinators to understand how the state assessment system works and how to interpret the results, particularly for open-ended items, but no direct training is given to teachers. Since 1993, information on open-ended items has been released for every grade level through sample papers and rubrics, with the goal of helping teachers understand the assessment and standards better. Assessment staff have just recently been charged to work with local school personnel to build assessment capacity. They will provide support to teachers for improving local assessments by incorporating current assessment tools, such as open-ended items, and by adding other types of assessments, such as portfolios.

Recent Forces for Change

Assessment is a high-profile item politically in North Carolina. Report cards have been instituted for districts to report results on end-of-grade tests, and a "state of the state" report is issued to compare North Carolina to the nation as a whole.

Accountability, in one form or another, has also been a buzzword in North Carolina recently, although the state has always emphasized public reporting of results. Since the 1980's, the state has been moving toward stronger and stronger accountability programs. The state board of education formed the ABC Program in response to a request by the state legislature to downsize the Department of Public Instruction and reorganize the public school system. The key features of the program are accountability for student achievement focused on the basics (reading, writing, and mathematics)

and local control of the public schools. This program will have financial rewards and sanctions that could involve the removal of principals or teachers at some point.[1]

A related, but somewhat different program is the governor's Standards and Accountability Commission, which is in its third year of existence. Its primary mission is to set standards for student achievement, and it is currently considering a broad-based assessment system that would strengthen local educators' abilities to determine levels of achievement. Assessment staff see the goals of the Standards and Accountability Commission as complementary to their own goals of improving instruction and ongoing assessment through their end-of-grade test system. In July, 1996, the commission will report to the state board, and their report could potentially change North Carolina's entire assessment program again.

Curriculum, Instruction, and Assessment

In North Carolina, 1991-92 was the last year of census norm-referenced testing. Assessment and curriculum/instructional staff worked in close collaboration to develop the current state assessment system of end-of-grade tests in reading, math, social studies, and science. In the words of Dr. Chris Averett, the assessment director, assessment and curriculum/ instructional staff were "joined at the hip" as they met in teams to work on items and decide on reading passages. During the 1992-93 school year, the end-of-grade tests were implemented in grades 3 through 8. They are given annually and are census tests. Both multiple-choice and open-ended items are included at grades 5 and 8, writing is tested at grades 4 and 7, and only multiple-choice items are utilized at grades 3, 4, 6, and 7. The tests are used primarily for accountability at the local school level; other uses include program improvement and documenting change.

Since the development of these tests, there has not been as much close contact between staffs, although assessment and curriculum specialists do work together on analysis and reporting tasks. Reading specialists also work closely with specialists in other curricular areas; for example, at the time of our interview, they were reviewing items for the grade 5 and 8 open-ended assessment.

North Carolina also has Benchmarks of Proficiency for reading and writing, for kindergarten through 12th grade. Benchmarks were developed in response to requests for more specificity in the curriculum framework and more guidance in interpreting assessment results.

The Influence of NAEP on Curriculum, Instruction, and Assessment

The primary influence of NAEP in North Carolina has been on the reading curriculum and framework. North Carolina's reading assessment is closely aligned with the framework, so it is hard to separate NAEP's influence on one from its influence on the other.

When the legislative mandate came through to develop the end-of-grade tests, the math curriculum had just been revised to meet the standards of NCTM, and the reading/language arts curriculum was due to be revised (as part of its regular cycle). As an agency, the department felt it was important to seriously evaluate the reading curriculum because they were planning to develop a new reading test that would be used for several years. North Carolina's goal for assessment and curriculum was that both be aligned with national standards. The department believed that the NAEP framework and assessment represented the standards for language arts more clearly than anything else available at the time.

---

[1] ABC= A for Accountability, B for high standards in Basic areas, and C for local Control or flexibility.

In the summer of 1990, at about the same time the framework for the NAEP reading assessment was being developed, then Assessment Director Bill Brown and Communication Skills Chief Consultant Cindy Heuts invited experts from around the United States to come to North Carolina and undertake the revision of the curriculum and assessment. Staff in North Carolina were in close touch with NAEP developers during this period, often obtaining NAEP materials as soon as they were developed. In this way, NAEP heavily influenced both North Carolina's framework and assessment. NAEP served to reinforce changes that were already underway and was also useful for informing local staff about planned or needed changes.

The resulting language arts curriculum framework comprises reading, writing, speaking, listening, and viewing. Whereas the former framework viewed reading as primarily skills-based, the new framework is organized around four broad goals. According to Dr. Averett, the new framework is very balanced in its approach to reading, reflecting a holistic model which addresses phonics as one of the cueing systems. The holistic, constructivist philosophy reflected in the language arts curriculum is also evident in the redevelopment of reading competencies for teacher education for kindergarten through grade 12.

Drafts of the framework were sent out all over the state for review; input was solicited from teachers, teacher educators, supervisors, and superintendents. After feedback was incorporated, the resulting framework was presented to the state board for approval, which was granted in February, 1992. The Division of Curriculum and Instruction subsequently took responsibility for disseminating the curriculum framework around the state. No changes have been made to the framework since its adoption.

According to Dr. Mary Rose, English language arts consultant, teachers are keenly aware that the curriculum framework and tests are very closely aligned. Presumably, these teachers also understand that students who learn material covered on the framework also perform well on the end-of-grade tests. Although one would be hard pressed to say that all reading instruction is carried out exactly the same way, because teachers do have decision making authority in their classrooms, this would seem to indicate that reading instruction in North Carolina is strongly influenced by NAEP.

The end-of-grade assessment requires students to read and write essays about authentic passages. Teachers participated in the holistic scoring of these essays during summer workshops, and Dr. Rose indicated that many teachers felt this was the best staff development they had ever experienced. Although this program was subsequently cut, Dr. Rose felt that the scoring workshops had a long-term, positive impact on both the teachers and students. Students will continue to exercise higher-order thinking skills in their learning, and teachers will be able to use what they learned in their scoring workshops to help their students.

Finally, Dr. Averett noted that when the curriculum was first revised, staff in the communications skills area of the department held workshops across the state to familiarize teachers with this new approach. She felt that these efforts really paid off because scores on the TSA reading assessment actually increased between 1992 and 1994 in North Carolina.

In addition to its influence on the reading curriculum and assessment, NAEP had an impact on public opinion and on perceptions of the school system. When North Carolina gave up its norm-referenced test, NAEP became the primary vehicle for national comparisons. The release of NAEP results is a big event in the state, and the release of local scores is likewise.

Limitations of the TSA

Lag time in reporting, lack of district results, an unpredictable assessment schedule, and a lack of understanding of the achievement levels were problems that North Carolina cited regarding the TSA program. It is worth noting that North Carolina linked its own 8th-grade mathematics assessment to the TSA in 1992, and have been projecting results onto the NAEP scale for both the state and districts since then. Dr. Averett felt this was an extremely powerful tool for district-to-state comparisons. In addition to providing district comparisons, the linkage also allowed North Carolina

BEST COPY AVAILABLE

to continue its trend line in mathematics through 1994, which it had counted on doing before the program funding was canceled.

Overall Evaluation of TSA Program

North Carolina is a strong supporter of the NAEP TSA program. According to Assessment Director Chris Averett, North Carolina will continue to participate in the TSA because of the necessity for national comparisons.

Dr.Rose felt that information from the NAEP program has been shared among educators in the state, but it has not necessarily been "taken to heart" as much as it should have been. She felt that state assessment results have garnered much more attention because these results are used to hold schools accountable; e.g., they are used to determine merit pay and bonuses. In this regard, the state assessment has much more of an impact in North Carolina than does the TSA.

Overall, the TSA program is seen as having a positive impact in the state, and both Dr. Averett and Dr. Rose felt that participation will continue to be worth the time, effort, and money spent.

Case Study for Pennsylvania

Overview

Pennsylvania has a diverse and large school population, ranging from students in small rural schools to large, inner-city schools. In spite of such diversity, Pennsylvania has taken an active role in developing statewide content area frameworks and sophisticated state assessments in reading, math, and writing.

Assessment activity in Pennsylvania is high profile. The Pennsylvania state assessment is currently favored for school accountability, even though strong local control is key to Pennsylvania's education system.

The role of NAEP in this complex and evolving education system is not predominant. With expanded state assessments that are aligned with instruction and that can be reported at district, school, and student levels, Pennsylvania district and school-level staff are becoming less and less motivated to take the time to participate in the NAEP TSA.

The State of Education in Pennsylvania

Organization

The Pennsylvania Department of Education consists of around 1200 employees. This state is known for having one the smallest numbers of state employees per capita in the United States. The department is divided into two units: (1) elementary and secondary, the larger of the two units, and (2) post-secondary/higher education. The Division of Evaluation and Reports is part of the elementary and secondary unit and is housed within the Bureau of Curriculum and Academic Services, to tie together curriculum areas with assessment. The Division of Evaluation and Reports has a number of functions, which include designing and implementing the state assessment program; providing assessment results to districts, state policy-makers, and the general public; and providing staff development to teach educators how to administer the statewide assessment performance tasks. Staff in the Division of Evaluation and Reports provide training to teachers and other district and school employees on how to understand the relationship between aspects of instruction and tasks, and items on the state assessment, and to utilize and interpret the state assessment results. In addition to these tasks, the division is responsible for administering the state NAEP.

The Curriculum division works closely with the Evaluation and Reports division, and is charged with the development and implementation of state content frameworks and their associated standards.

Recent Forces for Change

Pennsylvania has undertaken an education reform effort that involves implementing a set of regulations called Chapter 5, consisting of 53 student learning outcomes in nine goal areas and increased graduation requirements. All districts are required to develop strategic plans to implement Chapter 5. In addition, Pennsylvania receives some Office of Educational Research and Improvement money to work on the integration of science and the arts. Aspects of Pennsylvania's assessment system are also likely to be affected by broader policy changes attendant upon the election of a new Republican governor in 1995.

Curriculum and Instruction

A set of reading outcomes are part of Chapter 5 and these are consistent with the state reading framework. Even though there is a state reading framework, there is no state-mandated curriculum. Districts can meet the reading outcomes using textbooks, phonics, whole language, or any other method or approach they select.

Assessment

The Pennsylvania system of state assessment includes tests in math and reading, as well as a writing test. Assessments in science, arts, and social studies are currently being considered. Beginning in 1995, all students in grades 5, 8, and 11 have been tested in math and reading. Half of the students in grades 6 and 9 are currently assessed in writing.

The reading, math, and writing assessments are reviewed and revised on a yearly basis by a group of teachers from across the state, with assistance from outside contractors as needed. The current math and reading assessments include approximately 200 multiple-choice items and seven open-ended tasks per student—some items are taken by all students and some are matrix sampled. The writing test uses the same prompts for both 6th and 9th graders, and scorers are not told which grade they are scoring. This method is used in order to create a more uniform scale of writing performance for the scorers. One of the biggest benefits of the writing assessment, which began in 1991, has been the training that teachers have received in how to score the assessments. Student writing scores have been improving.

These reading and math assessments are aimed at program improvement and planning, but now, with the release of school scores, they have also begun to be used for accountability. Although all of the state tests have some effect at the local level, the state department of education must be very careful that the state assessments are not used in a manner that would upset the balance of local curriculum control, which is fundamental to Pennsylvania's education system.

In addition to the Pennsylvania state assessments, most districts, at their discretion, use some type of nationally normed test in grades K-12. Districts are also required to develop their own standards.

The Influence of NAEP on Curriculum, Instruction, and Assessment

The TSA has not had a significant direct influence on reading curriculum in Pennsylvania because the state reading framework was already well established by 1992. However, the NAEP reading framework was one of the many influences on the development of the Pennsylvania state reading assessment. Additional influences may emerge as state officials are now embarking on an integrated plan to set achievement levels, design content standards, then revise the state assessment, and finally adjust the achievement levels.

Pennsylvania has had a reading assessment for almost 30 years; the current assessment is influenced by the NAEP reading framework. A curriculum framework for reading, writing, and speaking across the curriculum, called the Pennsylvania Comprehensive Reading Program (PCRPI) was implemented in 1978. In 1988, Pennsylvania moved towards a whole language reading approach and designed the PCRPII. In 1991, the Pennsylvania assessment was revised to further reflect a whole language approach. The most recent changes included using full passages, having students respond to literature, adding performance tasks scored with a four-point rubric, and reducing the number of multiple-choice and summary questions. The NAEP reading stances are integrated into Pennsylvania's reading assessment rubrics. The NAEP stances have also influenced how teachers were trained to score the performance tasks.

The NAEP achievement levels have also had a significant influence on the Pennsylvania state assessment. In 1995, state assessment scores were reported on a quartile basis. In the future, Pennsylvania education officials hope to create performance levels similar to NAEP's. There is a strong interest in standards-based reporting in Pennsylvania. Education officials plan to set achievement levels for Pennsylvania in time for the reporting of the 1996 state

BEST COPY AVAILABLE

assessment results. They have hired a contractor to help them go through a process similar to the one undertaken to establish the NAEP achievement levels.

Limitations of the TSA

The TSA poses a number of minor problems for the state of Pennsylvania. First, it competes for student and staff time with the state's assessment program, as well as the district assessments. As a result, Pennsylvania had to drop the 8th-grade NAEP for 1996. Pennsylvania cannot mandate participation in NAEP because it would be anti-local control. It is particularly difficult to justify participation in the TSA to districts and schools when they do not receive any results that relate directly to them. Also, some parent advocates oppose the background questions asked on the TSA, particularly questions about family. Pennsylvania will not require students to answer the NAEP background questions.

Overall Evaluation of the TSA Program

Overall the NAEP has had a mixed influence on Pennsylvania. While the NAEP framework has affected the design of the Pennsylvania state reading assessment, the TSA has had difficulty competing with the variety of assessments on both the state and local levels. TSA results have had some influence at the state level, where policy makers use them to encourage reform. Generally, Pennsylvania education officials think it is worth continuing with state NAEP as an overall indicator of how Pennsylvania is doing in comparison to the rest of the nation.

Case Study for Rhode Island

Overview

Many changes in curriculum, instruction, and assessment have occurred recently in Rhode Island. The NAEP TSA has had a minor influence on some of these efforts, including the development of the English/language arts curriculum framework and of performance assessments in reading. Most important to Rhode Island, NAEP confirms that the state's recent efforts to improve the type and quality of instruction in the classroom are in concert with national efforts. Furthermore, the improved performance in reading in 1994 confirms that Rhode Island's major initiative, involving reading at the early grades, is working.

The State of Education in Rhode Island

Organization

Rhode Island is a local control state with 36 districts, and it serves 145,000 students in grades K-12. It has a commissioner of education who is appointed by the Board of Regents; the Board of Regents is appointed to staggered terms by the Governor.

Curriculum and instruction responsibilities at the state level reside in the Office of Instruction. Director Marie DiBiasio has a staff of three consultants, who have expertise in the areas of mathematics, science, and early childhood education. Ms. DiBiasio's area of expertise is English/language arts. These four areas have been the major foci in Rhode Island over the past few years. Though each staff member has an area of expertise, all four serve as generalists more than specialists, and they are expected to serve as district resources in various aspects of curriculum and instruction.

The Office of Assessment is responsible for overseeing state assessment; conducting program evaluations; providing enrollment projections and other analyses for the commissioner of education; and managing the department of education's Management Information System (MIS). The director of the Office of Assessment is Dr. Pat DeVito. The assessment section has a staff of 10; the MIS section, 8.

Recent Forces for Change

In 1987, the state legislature passed the Rhode Island Literacy and Dropout Prevention Act, a major initiative that impacted curriculum and instruction in the area of early childhood education. The focus of the legislation was on providing a high quality early childhood program, particularly in grades K-3, as a means of improving literacy and preventing dropouts. This initiative included a reading component that promoted integrated language arts, use of literature beyond the basal reader, and reading and writing as process. Dr. DiBiasio was hired to work with the districts to implement the legislation. As the representative of the department, she was responsible for getting districts to understand the current research and practice in reading so that district curriculum and instructional programs could be aligned with the mandate. As a result of the Literacy and Dropout Prevention Act, English/language arts and mathematics standards for grade 3 were also developed.

In the years since the Literacy and Dropout Prevention Act was enacted, Dr. DiBiasio has seen changes in many classrooms across Rhode Island. She reports that the whole language concept is being used in many districts, and writing and literature reading is occurring in many more classrooms. Unfortunately, in some classrooms the whole language concept is misused, and basic skills and phonics have been skipped, to the detriment of student learning. According to Dr. DiBiasio, these problems underscore the need for continuing professional development, not only in the delivery of instruction but in related areas such as authentic assessment.

Recently, the Office of Instruction has been unable to provide the professional development assistance that districts need to continue reforming their programs. When the Act was passed, a significant amount of money was set aside for its implementation, and funds were increased in the first two-to-three years. The intention was to level off funding at a reasonable level, but resources have since become tight in Rhode Island. The amount of funding currently appropriated to continue teachers' professional development is inadequate, according to Dr. DiBiasio. This does not mean that the efforts have stopped at the district level, but they have been slowed down.

Curriculum and Instruction

The state has curriculum frameworks for grades K-12 which districts may use in developing district curriculum guides and instructional programs. Although curriculum standards are included in the frameworks, the standards are not mandated. The department intends for the statewide assessments in the relevant content areas to be closely aligned with the frameworks. A framework for health education, first developed in 1987, is presently being revised. Mathematics and science frameworks have recently been completed, and a draft of the English/language arts frameworks, started in January 1995, is currently being reviewed and is expected to be approved by June 1996. Dr. DiBiasio led the development of the English/language arts framework and worked with a cross-section of K-12 teachers and district staff in the reading/literacy/writing community.

Rhode Island's Basic Educational Program (BEP) requires districts to self-monitor their district plans, which include curriculum guides, programs of instructional strategies, materials, and processes for evaluation in each content area. The department's responsibility is limited to a review of these plans.

Assessment

State law requires statewide testing but does not specify grades and subject areas except for health education. Historically, Rhode Island has conducted statewide assessment in physical fitness, health education, reading, writing, and mathematics at grades 4, 8, and 10. Because of a lack of funding, physical fitness is no longer assessed. Health knowledge is still assessed by state law. Reading and mathematics are assessed through upper-level subtests of MAT, version 7 (MAT7). The MAT7 reading and mathematics subtests used assess reading comprehension and mathematical problem solving, respectively.

An authentic writing assessment has also been administered annually to all students in selected grades since 1986-87. In 1994-95, grades 4 and 8 completed the writing assessment. The writing assessment is conducted over a two-day period and follows the process approach to writing. Students are a given a prompt; they write down key words and topics to think about; and they have 45 minutes to write a draft essay. The next day they get back their rough work and revise it. This revised version is holistically scored by Rhode Island teachers for the assessment.

According to Assessment Office Director Dr. Pat DeVito, Rhode Island is presently on track with a five-year assessment plan that will take them to 1999-2000. The focus of the plan is an increasing reliance on performance assessment in a variety of content areas including reading, writing, mathematics, science, and health. This year, 1995-96, the state will expand its writing assessment program to include grade 10 along with grades 4 and 8; continue to assess at grades 4, 8, and 10 with MAT7; and institute an on-demand performance assessment in mathematics and health at grade 4. The state also hopes to develop and pilot performance assessments in reading. An outside contractor has been hired to help develop more performance assessments. Rhode Island's goal is full implementation of on-demand performance assessment in reading, writing, mathematics, science, and health in grades 4, 8, and 10 by the year 2000.

One of the major thrusts of Rhode Island's assessment program has been "total inclusion," and Rhode Island has received national recognition for this effort. In the past, the state had exemption policies that allowed students to be excluded from testing if they were in special education classes 50 percent or more of the time or in a limited English proficient program. Total inclusion was field tested in Rhode Island in 1994-95, and it is being implemented in 1995-96. The accommodations made for total inclusion are varied, and they have not been as costly as originally expected. Dr. DeVito indicates that the formerly excluded students are being instructed every day in classrooms; the

accommodations made for these students in the assessment administration, therefore, have mainly been those made to facilitate testing the students. Examples include permitting the use of computers in the writing assessment, allowing additional time for completing assessments, reading directions orally, and re-reading directions or questions for clarity and understanding.

Rhode Island is a partner in two assessment-related initiatives—New Standards and State Collaboration on Assessment and Student Standards (SCASS) Projects of the Council of Chief State School Officers—and both have helped the state develop its assessment program. Rhode Island's Goals 2000 Panel is pushing curriculum frameworks and performance assessment. The Office of Assessment staff has worked with this Panel, but most of the work and 90 percent of the Goals 2000 funds are focused at the district level. Assessment and accountability are hot political issues for the commissioner, the Board of Regents, and the department, but less so for the legislature, according to Dr. DeVito. The legislature wants to see assessment results and wants accountability; however, it has not provided sufficient funds for a comprehensive system.

Influence of NAEP TSA on Curriculum, Instruction, and Assessment

Rhode Island has participated in all administrations of the NAEP TSA, and minor influence on reading curriculum and instruction was exerted, mainly because the TSA came after passage of the Rhode Island Literacy and Dropout Prevention Act. Because of the act, major efforts in the area of reading instruction, NAEP-like in content, were already underway when the TSA began. Dr. DiBiasio brought some of NAEP's influence to the development of Rhode Island's English/language arts curriculum framework through her participation in the development of NAEP items and her understanding of the framework. According to her, however, the greatest impact of the NAEP TSA has resulted from the 1994 reading results, which provided evidence for her and others in Rhode Island that what they were doing in the area of English/language arts instruction appeared to be working. The 4th grade that took the 1994 NAEP TSA was the first cohort to go through grades K-3 after implementation of the Literacy Act, and their scores were higher than those of Rhode Island students in previous samples. The results from the state writing assessment also reflected an improvement in performance.

In assessment, NAEP has had a minor influence, according to Dr. DeVito. As with curriculum and instruction, assessment efforts in Rhode Island were already undergoing changes and were subject to numerous influences. The NAEP reading framework, the format of the assessment, and the types of items in the assessment influenced the look of the state's reading assessment to some degree. NAEP had some impact on increasing the emphasis on literature and higher-order thinking skills, both in the development of the new performance assessments and in the choice of the norm-referenced test that Rhode Island administers. NAEP has also influenced the state to include more performance-type assessment, use authentic passages and constructed-response items, and include greater numbers of students with disabilities and second language learners. According to Dr. DeVito, NAEP is a well-respected national assessment and a scientifically rigid system that allows Rhode Island to say, "It's going the same way we are talking about so we are not out on a limb."

NAEP is one of the components of the five-year assessment program proposed for the state. The program includes portfolios at the local level, on-demand performance assessment, the MAT7 at the state level, and biennial testing in NAEP.

Limitations of the NAEP TSA

Despite the progress made in classrooms across Rhode Island as a result of the efforts launched by the Literacy and Dropout Prevention Act, many of Rhode Island's classroom practices do not reflect what is most current in reading research. Therefore, one of the factors that has limited NAEP's utility in Rhode Island is its lack of alignment with classroom practice.

Dr. DeVito says that the lack of local results also limits NAEP's utility to districts in Rhode Island and in turn makes it more difficult for the state to recruit schools for NAEP participation. Donating students' time without a return of school or district results is a difficult thing for locals to do. Furthermore, the lag time in reporting is a major factor limiting utility of NAEP results. States and districts do not want to wait a year or more for the results of the assessment.

Finally, Dr. DeVito feels that the unpredictability of the assessment schedule is a problem. Not knowing which content areas and grades will be included in the assessment until close to assessment time makes it extremely difficult to recruit schools. Dr. DeVito indicated that the Education Information Advisory Committee and the assessment directors have suggested that NAEP's National Assessment Governing Board formulate contingency plans that depend on different levels of funding from Congress. Contingency plans would give states and local staff a better idea of the assessment schedule, allowing them a reasonable time to decide whether to participate.

Overall Evaluation of the TSA Program

The NAEP TSA has had a "generally positive" impact to date on education in Rhode Island. Despite this, and the fact that the state has included NAEP in its five-year assessment plan, Rhode Island's participation in future NAEP TSAs is questionable. Soliciting districts for the 1996 administration of NAEP has been difficult. Because of Rhode Island's size and the requirement to provide a sample of 2,500 students, the burden on schools has been greater in Rhode Island than in most other states. For example, the sample of 2,500 8th graders would include 23 percent of Rhode Islands' 8th-grade population. Consequently, many schools must participate at each administration. With the problem of overburden coupled with Rhode Island's own development of performance assessments and the state's participation in efforts such as New Standards, which has developed referencing exams in mathematics and reading and may be considered an overlap with NAEP in mathematics and reading assessment, the uncertainty of Rhode Island's future participation in NAEP grows.

Case Study for West Virginia

Overview

West Virginia's legislature mandates participation in NAEP, primarily to enable educators to compare the performance of the West Virginia students to that of other students in the nation. Although the mandate for participation has made educators across the state aware of the existence of the NAEP program, NAEP has had limited impact on instruction because of local control of curriculum and instruction. In areas where awareness of NAEP has peaked, educators have increased efforts to incorporate the teaching of higher-order thinking skills into reading instruction.

NAEP has had *some* impact in West Virginia with respect to the content frameworks. Personnel in the state department of education expect that NAEP's influence will increase in the future, particularly after the present legislative session, when the current reading framework and testing program are to be revised.

The State of Education in West Virginia

Organization

West Virginia's education system is structured by county, and each county is a district. The Office of Instructional Services was recently reorganized and has a staff of 12 full-time professionals. One is Melvin Graham, who became the Title I K-12 reading specialist in the past year. Mr. Graham's primary duties are working with Title I and providing technical assistance to local education agencies; other duties include working with the reading/language arts supervisor to develop and disseminate the language arts curriculum. The technical assistance includes conducting workshops, conducting regional meetings to provide information to local staff, and providing assistance to low-scoring schools that have been targeted for program improvement. Graham's staff provide teacher in-service training upon request, but they have no responsibilities for teacher pre-service training.

Assessment activities are located within the Office of Student Services and Assessment, which oversees the assessment division as well as a throng of student programs (e.g., Drug Free Schools, Dropout Prevention). Karen Nicholson, the assessment director who was interviewed for this study, is the assistant director of this Office. She works with a coordinator and has two technical (non-professional) staff who help with the State/County Testing Program. Ms. Nicholson and her coordinator also manage all the field activities associated with the Statewide Testing for Educational Progress (STEP) and NAEP programs. (The State/County Testing and STEP Programs are described below.) Assessment is a high-profile activity in West Virginia. Although it is largely the responsibility of the department of education, the STEP Program is controlled by the legislature, and the department is required by law to make assessment results public. About 60 percent of the decisions regarding assessment are made in-house, while the rest are shared with county personnel, the state assessment advisory committee, and others.

Recent Forces for Change

A number of education initiatives are currently in progress. West Virginia is participating in Goals 2000, but the program has not had much effect yet. There is a strong emphasis on community team building, which involves partnerships between business leaders and educators. Site-based management has been a potent force in the department, and it is responsible for the preeminence of the charge for technical assistance to local schools and school systems. The Governor has fostered a number of literacy programs.

Some local incidents have occurred to encourage reading instruction to go in one direction versus another (e.g., phonics versus whole language), but none of these incidents has had much lasting power. One of the biggest controversies has concerned objections of the political right to particular aspects of textbooks used in schools. On the

whole, achievement levels in reading have been good, and this has deflected attention away from reading and toward areas like mathematics, in which West Virginia students have not performed as well as in reading.

Reading Curriculum and Instruction

West Virginia has a language arts curriculum which establishes a framework of instructional goals and objectives that must be delivered in all public schools in the state. The framework comprises four traditional areas—reading, listening, speaking, and writing — and includes a new area, viewing, which is intended to address discourse regarding drama, theater, film, television, and computer technology. The reading curriculum framework is fairly general, and LEAs tailor the framework's guidelines to their own needs. As a result, some school districts have modernized their reading instruction delivery a great deal, whereas others maintain a textbook-driven approach.

Assessment Program

West Virginia has separate frameworks for curriculum/instruction and assessment, and current revisions of the two frameworks are expected to bring them more closely into alignment. West Virginia's assessment program has two components: the State/County Testing Program and the West Virginia STEP. Both programs operate on an annual basis, and all students are required to participate in them.

The State/County Testing Program administers a norm-referenced test (CTB-McGraw Hill's California Test of Basic Skills, fourth edition) to students in grades 3, 6, 9, and 11, and conducts a writing assessment at grades 8 and 10. The State/County Testing Program is used to provide accountability, to determine accreditation, and, in some cases, to make decisions about individual students.

The STEP Program includes mandatory participation in NAEP and the use of criterion-referenced tests of reading, mathematics, and composition to assess students in grades 1 through 8. Items for the STEP tests are generated by a contractor from a set of specifications provided by the department; teachers then meet to choose items that reflect current curricular and instructional practices for the assessment. Teachers also determine the cutscores for mastery on the assessment, and they holistically score the writing composition assessment. The STEP criterion-referenced reading test for grade 4 includes sections for listening, reading comprehension using authentic passages, and writing; multiple-choice, short answer, and extended-response questions are included on the assessment. Actual literary passages were not included in the previous assessments; these are new to the assessment. The criterion-referenced tests are aligned with instruction, and they were revised during the 1994-95 school year.

The West Virginia legislature met in mid-January, 1996 to decide the future of the STEP Program. The legislature, in February, changed the language of the law. The STEP Program is no longer mandated. Counties and schools have the option to give it if they choose. There had been some pressure to drop the criterion-referenced portion of the STEP test because of the time and resource commitments it requires.

Influence of the NAEP TSA on Curriculum, Instruction, and Assessment

According to Melvin Graham, reading curriculum specialist, the reading framework was developed by a committee of practitioners with NAEP in mind. The most recent revision of the curriculum reflects new research on reading; further development is expected to occur in spring, 1996, after the legislature decides the future of West Virginia's testing program. Some districts have made extensive changes in reading instruction in past years, incorporating the teaching of higher-order thinking and other skills into their curricula; other districts are still very textbook driven. The state superintendent has placed a high emphasis on the NAEP assessment.

Assessment Director Nicholson reported that while developing their assessment, department staff were aware of the NAEP frameworks and of the move toward more short-answer and extended-response writing. The assessment director and reading specialist were careful to emphasize, however, that West Virginia local education agencies did not buy into

NAEP in a wholesale fashion. Rather, NAEP was examined critically in the context of changes in the field of reading, and a decision was made to move in the same direction. For example, although the reading community may have bought into higher-order thinking skills, other policy makers do not always have knowledge about recent developments in reading to make informed decisions about including higher-order thinking skills in state assessments. Once the policy maker sees this type of skill being assessed in NAEP, however, he or she understands the need for this type of change to occur. In fact, the appearance of extended-response questions on the NAEP assessment paved the way for these types of items on West Virginia's own assessment. Therefore, the TSA served to reinforce the validity of changes that were already underway in the reading field and legitimated these changes in the eyes of many.

As stated above, West Virginia law requires NAEP participation, and NAEP is the only vehicle for comparing the state's schools to the nation as a whole. NAEP helps provide the justification for doing additional assessments, selecting different assessments, or changing the system. In addition, West Virginia participates in activities with the Southern Regional Education Board, and NAEP provides the common linkage for these activities.

West Virginia looks at specific items on the TSA for information purposes; for example, the question about how reading performance varies by student background, choice of reading materials, and so on, but does not use these results specifically for planning. NAEP information is shared with county and district personnel, but decisions about what to do with it are made locally.

In the past, there has been little to no coordination between instructional and assessment staff in test development and other activities, but cooperation has increased recently. It will probably continue to increase in accordance with the new test adoption.

Limitations of the TSA

According to Melvin Graham, NAEP has had its greatest influence in areas where people know about it. In those areas of West Virginia where NAEP has been able to establish a presence for itself, there has been a greater tendency toward change and more acceptance of authentic types of assessment. Mr. Graham predicted that NAEP would have a larger influence in West Virginia in upcoming years because it exemplifies current thinking in the reading field and provides a good framework for instruction.

Overall Evaluation of TSA Program

West Virginia seems fairly satisfied with the NAEP program. There are calls for district, school, and student-level results, but these are not overly strident. Assessment staff are looking more closely at NAEP than ever before, primarily because they want students to have hands-on experience in mathematics and in science, and NAEP provides a good model for this type of assessment.

Case Study for Wisconsin

Overview

Students in Wisconsin are expected to perform well on national measures of academic achievement, and they have generally done so. Consequently, although Wisconsin participated in the 1992 and 1994 NAEP TSAs in reading and views the NAEP program positively, the influence of the NAEP TSA on the state's education program has been minimal. The performance of Wisconsin's students on the NAEP in reading has provided Wisconsin with confirmation that their language arts program is on the right path and that it is having a positive impact on student learning.

State of Education in Wisconsin

Recent Political Issues

At the moment, the most important political issue related to education in Wisconsin is a struggle that involves a change in the Department of Public Instruction (DPI). The legislature, basically at the prodding of the Governor, recently passed a bill, which mandated, as of January 1, 1996, a change in the name of the Department of Public Instruction to the department of education and the replacement of an elected state superintendent of education by a secretary of education appointed by the Governor. In December 1995, the Wisconsin Supreme Court placed a "restraining order" on the bill's implementation pending the outcome of a constitutional challenge. The bill, nevertheless, has created a state of uncertainty in the DPI.

Darwin Kaufman, the Director of the DPI's Office of Accountability, indicated that, should the constitutionality of the legislation be upheld, it is difficult to know what effects it would have on the education system in Wisconsin. An educated guess is a possible diminution of influence of groups such as the teachers' union and other curriculum and instruction organizations, which through the years have acquired tremendous influence in the education system. An appointed secretary would probably create an education program that reflects the Governor's agenda to a greater extent. However, the Governor and the department agree on many initiatives, including an increased attention on statewide standards and accountability.

Organization

Presently, curriculum and instruction responsibilities in the DPI reside with curriculum consultants—one in each content area. Jacque Karbon, the reading curriculum consultant, does not have a staff that supports her although she works in collaboration with other curriculum specialists, Office of Accountability staff, and other program staff. Her responsibilities include being a resource for local districts and a liaison between the state department and local districts. She keeps districts informed about current research and about state and federal programs and legislative requirements; serves as the state liaison to the Wisconsin state reading association, participates in other professional organizations such as the International Reading Organization, and helps districts network with each other.

Statewide assessment responsibilities reside in the Office of Accountability. The director has a staff of 12 who spend about 75 percent to 80 percent of their time directly on the statewide assessment program. Staff members also work closely with other staff in the DPI. For example, the reading curriculum consultant spends about 15 percent to 20 percent of her time working on reading assessment, and the Office of Accountability has one staff person spend about 30 percent to 40 percent of her time with the Title I program.

Wisconsin has a long tradition of local autonomy. There are over 400 districts in the state. There is also a system of 12 regional offices, called Cooperative Educational Service Areas, that are not part of the state DPI but provide

services to "member" districts. That is, the districts pay for the services provided by their Cooperative Educational Service Area.

Curriculum and Instruction

Wisconsin does not have a statewide curriculum, but districts are mandated to have a district curriculum. The state offers districts a series of Guides to Curriculum Planning in the different content areas that districts use at their discretion. These guides were revised most recently in 1986. Districts decide on their district goals and objectives; the areas of specific curriculum emphasis; time allocation across the curriculum; and materials to be used. Consequently, the content reading programs vary across districts. Some districts emphasize the constructivist view of reading with an integrated language arts program and a strong emphasis on the writing process using extensive children's literature, while other districts have a stronger emphasis on direct instruction with more use of basal readers. Generally, however, the districts in Wisconsin are open to what is current in the domain of reading as well as in related areas such as learning theory.

Wisconsin has a number of grants for projects that are expected to impact curriculum and instruction in Wisconsin. One is for a project called Connecting the Curriculum that is aimed at integrating curriculum and involves teachers in action research. Another grant is for the development of challenging content standards in the areas of arts, language arts, foreign language, and social studies. Wisconsin proposed to develop content standards in each individual area as well as pieces that connect all four areas together. Wisconsin also has a grant for developing frameworks in the areas of mathematics and science.

There is also a movement in Wisconsin to more definitively support phonics in reading instruction. Although phonics has always had a place in the reading curriculum in Wisconsin, recent efforts have been made to be more explicit about teacher training in phonics. A bill is also pending that would require teachers applying for certification or re-certification to show that they have successfully completed instruction in teaching phonics.

Assessment

Wisconsin has a statewide assessment program that was implemented for the first time in 1989. Prior to that, the state had a testing standard that required districts to administer a test of their choice. Until this school year, statewide testing was conducted in various content areas at grades 8 and 10 and in reading only at grade 3. Beginning this school year, 1995-96, testing will also be administered on a voluntary, trial basis at grade 4. The assessments are conducted annually and all students in the relevant grades participate.

During this 1995-96 school year, the students in grades 4, 8, and 10 will be tested in mathematics, science, social studies, reading, and written English. The assessment consists of a multiple-choice test in each subject area, and three short answer questions in each area except written English. In written English, two essay questions are administered at each of the grade levels. The tests were developed by the Psychological Corporation and based on two commercial products: the SAT, version 8, and Goals. Psychological Corporation is coming out this year with a new version of the SAT, version 9. Wisconsin is presently using the SAT 8.

The 3rd-grade reading test is a Wisconsin-developed product; the DPI develops a new reading test each year. The test consists of four passages with at least one passage expository and the remainder narratives.

Wisconsin has also been working with the University of Wisconsin, Madison, for three years on the development of performance assessments in the areas of mathematics, language arts, and science. The funding for that project was surprisingly cut last year as a result of a drive by a group of people opposing performance assessments. There is interest in restoring the funding, and there is support for this from the Governor as well as others in the education community.

Performance Standards

Presently, Wisconsin has no statewide performance standards, but has a statute that requires the state superintendent to identify low-performing districts and schools and to set state standards. The legislation does not identify how this should be done. This legislation is now being used as the motivation behind a recent effort to establish performance standards in the five content areas that are presently part of Wisconsin's assessment program. Work on performance standards is expected to begin this spring.

Because of the lack of statewide content standards and strong local control, the curriculum among districts, and sometimes even within districts, varies. This results in a problem that the state has been struggling with—what to base the content of their statewide assessment on. However, the state is making progress in moving forward a little more rapidly in the development of content standards, particularly with the recent grant to establish challenging content standards in language arts, humanities, and social studies. Wisconsin anticipates the articulation of the assessment with curriculum content to be easier once content standards are developed.

Influence of NAEP TSA

Curriculum and Instruction

NAEP has had a minor influence on reading curriculum and instruction in Wisconsin. It reinforced the validity of what is already underway in reading instruction. Feedback from teachers indicates that the nature of the Wisconsin 3rd-grade reading assessment has changed the way reading is being taught at the primary level and because the 3rd-grade reading assessment is influenced by NAEP, indirectly, NAEP is influencing how reading is taught in the classrooms.

NAEP documents, including the reading framework and the assessment format and items, have been used in promoting assessment literacy. Jacque Karbon has indicated that by comparing Wisconsin's 3rd-grade test with NAEP in reading, people understand better that there are different types of tests for different purposes. Local districts see Wisconsin's NAEP performance as confirmation that what they are doing in reading is working.

NAEP documents have also been used as resources for districts. One example relayed by Karbon involves one of Wisconsin's larger urban districts. The district was interested in developing a survey of its reading program. Jacque suggested that it look at the NAEP questionnaires from the NAEP report on reading literacy for examples of questions, how to collect information, how to formulate questions, what questions to ask, and how to report the information.

Assessment

NAEP's influence on assessment in Wisconsin has also been minor. It serves as a source of information as Wisconsin annually develops its 3rd-grade reading test. For example, longer reading passages used in the NAEP TSA confirmed and reinforced the use of such passages in the 3rd-grade reading test.

Kaufman indicated that the impact of NAEP TSA has occurred mainly by means of those people who participated in the development of the NAEP framework and assessment and who also work in the areas of curriculum, instruction, and assessment in Wisconsin. By participating in the review and critique of the NAEP frameworks and of assessment items, these people became familiar with and understood better the form of the assessment and the types of items being developed. By sitting down with the items and examining the results of the field testing, people increased their knowledge and ability to use the knowledge in the development of Wisconsin's 3rd-grade reading assessment. There is an expectation in Wisconsin that students will perform well on NAEP and they do. Living up to these expectations means that NAEP performance results are less a motivating factor for change than they would probably be if the results indicated poor performance.

Limitations of the NAEP TSA

Although both Karbon and Kaufman have indicated that the NAEP TSA has limitations that prevent it from being more influential in Wisconsin than it could be, they understand the limitations and frankly are not always convinced that there are acceptable changes that can be made to eliminate these limitations. Both agree that NAEP results are not sufficiently helpful for diagnosing instructional problems. Karbon points out, however, that NAEP results are useful in broad program planning.

As with other states, Wisconsin also finds the lag time to reporting a limitation. Kaufman explains that states put in a tremendous amount of time and energy in the few months before the assessment, recruiting schools and conducting the assessment. However, because of the long lag time, people have forgotten about the assessment by the time the results are presented. Karbon sees a positive aspect to the lag time states and locals must endure. She indicated that the lag time provides a good message to consumers who always put a tremendous amount of pressure for immediate feedback of test results. This message is that results of quality and utility in a complex assessment program require careful analysis that tend to take a substantial amount of time.

Kaufman also sees the fact that NAEP provides no district-level results as limiting its utility for districts. He indicated that it would also be easier to get locals to pay attention to the results if they were provided locally.

The unpredictability of the assessment schedule was seen as a problem not only in limiting its use but also for recruiting schools. Locals would be able to better plan for their district and school testing programs if they knew which areas and which grades are to be tested far in advance of the administration date. NAEP results would be able to play a more major role if districts and schools could be assured that the needed information would be available. The unpredictable assessment schedule has also made it more difficult for the state to recruit schools. If schools don't participate, NAEP loses most, if not all, influence in the district. In order to get cooperation of districts and develop interest in the results, the state must be able to communicate to districts and schools on an ongoing basis regarding the assessment results (grade levels and content areas) locals can expect to have for their use.

Finally, Karbon felt that the fact that the between-state comparisons make limited allowance for factors beyond educators' control limits the utility of the data. She believes that NAEP reporting must make the possible effect of these factors more explicit so that the influence of these contextual factors will not be questioned, thereby limiting the impact of the results.

Overall Evaluation of the TSA Program

NAEP TSA has had a "generally positive" impact to date on education in Wisconsin. The utility, however, of the information collected through NAEP has been limited by a number of factors, many of which, as has been pointed out either by Karbon or Kaufman, are not the fault of the assessment itself. In some cases, NAEP is being asked to be something other than it is intended to be—e.g., can NAEP really be used effectively for diagnosing instructional problems, and should it? In other cases, competing priorities hinder efforts to address limiting factors—e.g., decreasing the lag time substantially may affect the amount and quality of results reported.

Kaufman believes that NAEP is a fine program and that there is no other assessment program in the country that is more technically sound. However, he also believes that in general NAEP has tremendous potential that is not being realized. That is, the NAEP TSA collects a tremendous amount of data that could be useful to states, but states lack the capacity to use these data and to get information out to the people in the schools and to the public. Kaufman seeks the answer to how states can work with those at the federal level to figure out how to report information in ways that will make a difference and have people take notice given the states' and federal government's limited resources.

Case Study for Wyoming

Overview

Wyoming is a large state with a small, fairly homogenous population. With only 100,000 K-12 students in 400 schools in 49 school districts, this setting is ideal for testing and implementing the statewide school reforms underway since 1990. Wyoming has always been a local control state, and this is mirrored in the reforms now being implemented.

The NAEP TSA plays a useful role in Wyoming because there is no other statewide assessment and there are no plans to implement one. Consequently, as more and more decision-making power is handed down to the districts and schools, the TSA remains the only overall monitor of Wyoming's academic achievement.

The State of Education in Wyoming

Organization

The Wyoming Department of Education has a staff of 80-85 and is divided into three work groups. The first workgroup, Program and Learning, deals with federal programs, school accreditation, and health and safety. The second workgroup, Administration and Internal Operations, is responsible for administration and human resource development, and internal budgets and quality. The third workgroup, Support Programs and Quality Results, deals with school finance and personnel, data and technology, and outreach services (vision and hearing). The assessment director is housed in the Quality Results workgroup.

Recent Forces for Change

In 1990, the Wyoming state board of education and the state superintendent developed a school-based improvement program that features a school accreditation process. The program is now being implemented. Although it is not legislatively mandated, all districts in the state are participating in the accreditation process. Districts have until 1997 to comply with the state accreditation guidelines, which stipulate that districts and schools must develop a system of performance standards in the major content areas, as well as a system of measuring performance. Principals, teachers, and community members are all required to participate in this effort and work together to develop a school improvement plan. As part of the accreditation process, 10 school districts are selected each year and are visited by teams of state employees and employees from other school districts, who look at data on student achievement and review school improvement plans. If a school is not complying with the program requirements at the time of the team's visit, they are given enforceable recommendations and a time line. The local educators and community members determine the specific ways in which the recommendations are to be met.

Another education effort taking place in Wyoming is the use of a $363,285 Goals 2000 grant award. Approximately $150,000 will be given to districts in a competitive grant process. Funds are being used to develop a statewide education technology plan, and establish resources and means for districts to implement the plan.

A final, and important, force for change is a recent ruling by the Wyoming Supreme Court that the state school finance system is unconstitutional and must be completely revised by July 1997. The case revolved around a complaint from large school districts that small districts were getting more money per student, and that larger districts were not being offered equitable funding. As a result, significant changes in school programs will probably occur in Wyoming over the next few years.

Curriculum and Instruction

Because of the heavy emphasis on developing content standards at the local level, Wyoming has no state-level content area standards. For the accreditation process, most districts and schools are using national standards, such as those developed by NCTM to develop performance standards. Upon request, the state department of education will help districts develop performance standards, and a number of districts have made use of this assistance.

Assessment

School districts are currently allowed to select and use any available test or combination of tests to assess their students. The state department of education is not involved in district testing at all; staff members do not recommend tests or monitor testing. Districts are using a variety of instruments, including criterion-referenced, norm-referenced, performance-based, authentic, and portfolio assessments. Scores are reported at the district, school, or individual student level. Most districts assess in the grades of their choice on a yearly basis. The results of these assessments are used for district, school, and student improvement. In the near future, state officials will examine the utility of each district's assessments for measuring its accreditation performance standards.

The Influence of NAEP on Curriculum, Instruction, and Assessment

NAEP has had no overt influence on curriculum or instruction in Wyoming. According to Dr. Al Sheinker, Wyoming assessment director, it is possible that in helping schools develop their performance standards, state curriculum or assessment specialists, who are familiar with NAEP, have slipped in some NAEP influence, but it would have been entirely incidental. At the local level, interest in NAEP is quite low, as NAEP results have very little effect on individual schools or districts.

NAEP has had little influence on assessment in Wyoming since no standardized assessments have been developed or revised in recent decades at the state or local level.

Limitations of the TSA

Dr. Sheinker sees two major limitations to the TSA. First, he would like to see the time lag between assessment and reporting reduced so that the results will have a greater, more immediate impact. Second, he believes the NAEP would be much more useful to districts and schools, for accountability and informative purposes, if it were reported at the district, school, and even individual student level.

Overall Evaluation of the TSA Program

In spite of its limitations, the NAEP TSA has had a positive effect on Wyoming overall, Dr. Sheinker says. Although it has not directly influenced curriculum, instruction, or assessment in Wyoming, it has filled a void created by the lack of a state assessment program. The NAEP TSA provides a means for the state to measure its academic achievement over the years in relation to itself and other participating states. In addition to helping Wyoming monitor its academic progress, participation in NAEP has always rewarded the state with a high ranking among the participating states. As a result, virtually no districts refuse to participate and there is a generally positive and appreciative attitude toward the role that NAEP plays in Wyoming.